Automated analysis and transcription of rhythm data and their use for composition

Boenn, Georg

Award date:
2011

Awarding institution:
University of Bath

Automated Analysis and Transcription of Rhythm Data and their Use for Composition

submitted by

Georg Boenn

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Computer Science

February 2011

To Daiva, the love of my life.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank Prof John ffitch for being an inspirational supervisor who introduced me to Hardy and Wright (1938) and the Farey Sequence. Many thanks also to his wife Audrey for proof-reading the thesis and for giving me valuable feedback on style, which is very important when writing in a foreign language. My second supervisor, Dr Archer Endrich, gave me detailed and precise feedback on my work, for which I am very grateful. Thanks also to my Dean, Prof Peter Robertson, and to my Head of Department, Michael Carklin for their continuing support, and thanks to my employer, the University of Glamorgan, for sponsoring my tuition fees. I am very thankful for my family's love, support and patience during the last five years. Finally, I would like to thank Prof Clarence Barlow whose enlightening seminars on Computer Music I attended at the Musikhochschule, Cologne, in 1990. His algorithmic composition software Autobusk (Barlow, 1984) was an inspiration for me during those years. The seeds came to fruition in the work that I am now able to present.

Summary

In this thesis we propose a new method for quantisation, tempo tracking and transcription of onset data from audio recordings without knowledge of the underlying score in advance. This method is particularly useful for the detection of agogics, ornaments and syncopated rhythms within a musical performance. Therefore we can see many applications of our program within the areas of musicology, composition and performance analysis. We review the existing research within the field and we show that a successful transcription is based on the detection of duration classes and the application of similarity measures within a given analysis window. It is also based on the Farey Sequence as the underlying grid combined with algorithms that measure a weighting of both the quantised inter-onset ratios as well as the weighting of a specific beat-structure used for the transcription of the quantised data. As a by-product to the quantisation towards the unknown score our method extracts the tempo changes and the beat structure of the performance. In addition to the rhythmic analysis, we would like to demonstrate the creative aspects and generative outcomes of rhythmic modelling that is based on the same core algorithms as the quantiser.

Nomenclature

ALSA	Advanced Linux Sound Architecture
BPM	Beats Per Minute
BWV	Index of Johann Sebasitan Bach's complete works, "Bach-Werke Verzeichnis"
CAC	Computer-Assisted Composition
CPN	Common Practice Notation
CUIDADO	Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online
DMRN	Digital Music Research Network
EAM	Electroacoustic Music
GTTM	Generative Theory of Tonal Music
HMM	Hidden Markov Model
IOI	Inter-Onset Interval
IRCAM	Institut de Recherche et Coordination Acoustique/Musique
JACK	JACK Audio Connection Kit
MAP	Maximum a posteriori
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MMH	Many Meters Hypothesis
MPEG	Moving Picture Experts Group
nPVI	normalised Pairwise Variability Index
PPQ	Pulse Per Quarter note

SMPTE	Society of Motion Picture and Television Engineers
VAMP	A plug-in format for audio feature extraction developed at Queen Mary, University of London

Chapter 1

Introduction

"A musically gifted man measures time by music." (Leoš Janáček)[1]

1.1 Musical Time and the Problem of Musical Form

At a time when composers, musicians and recording engineers have tools available that can control every digital sample of a sound-recording, there are still unresolved problems with musical time and musical rhythm perception on many different levels that have a direct impact on the work of composers, musicians and on every music listener. For example, exactly how humans recognise an implicit underlying stream of beats within a musical piece is still unknown and poses a hard problem for computer-based music recognition (Davies, 2007; Dixon, 2001; Goto, 2001). In the related field of automated transcription of rhythmic onset data from audio recordings, some progress has been made and the purpose of this thesis is to demonstrate our own research efforts and results achieved specifically in the area of onset quantisation and transcription of musical rhythms. On a larger scale, rhythm relates to musical form and the problem of form in relation to the material that composers arrange in their work presents a continuing challenge. From the point of view of the composer, this has to do with the simultaneous control of a multi-dimensional parameter-space; from the point of view of the performers and listeners the problem is to reduce the multi-dimensional parameter-movements into a perception of unity. Future music-psychological studies should focus on how exactly we correlate and integrate a diverse set of musical parameters that evolve and change continuously and simultaneously over time. We assert that the results presented here are a useful contribution to the discussions on how we perceive rhythms and musical timing expressed by performers. Although we are aware of multi-dimensionality in musical practice and cognition[2], due to time and space constraints, we needed to concentrate on rhythm analysis from onsets only and could not integrate other musical parameters.

[1] Quoted from Wingfield (1999, p.221)

[2] see for example the works on Auditory Scene Analysis by Bregman (1990) and Wang (2006)

1.2 Context of Research and Research Questions

In this thesis we present our own system for automated rhythm transcription from audio onsets and how it relates to previous works in the field.

Our work was driven by the need for a reliable software tool that would be capable of rendering scores from musical performances in such a way that almost no editing of the score was needed. Specifically, the program should find a truthful representation of the musical rhythms that have been performed. One possible user scenario could be that a spontaneously created and improvised piece of music is recorded via a MIDI keyboard and then transcribed into a musical notation, which can be regarded as a truthful representation of this particular instance of the new piece that was created ad hoc, in realtime, freely improvised or perhaps being prepared from previous sketches. To capture rhythms truthfully is the most challenging task here and we had to develop a general model of performed musical rhythms, together with a new computational method of representing and analysing such rhythms from various origins. In essence, the work is about a computational analysis of performed musical rhythms and to match the outcome with an appropriate form of rhythmic representation that could then be translated into a format for musical notation.

One has to ask under which conditions such a system of rhythm detection can be useful and what problems need to be addressed and solved throughout its development.

The work of a composer incorporates to capture musical experiences. It involves the arrangement, the transformation and transcription of an inner musical experience into a written form, i.e., a score, which can be handed-out and communicated to other musicians and performers in order for them to re-create this new musical experience. This special kind of transcription and communication with performers lies at the heart of a composer's work, so it is only reasonable to reflect our computational approach from the standpoint of the musician and composer and also to point to certain compositional applications of the algorithms that are in use.

In our work we concentrate on the analysis of musical rhythms. Pitch phenomena are not taken into account, although a final software solution for transcription would of course need to include pitch data. While it is trivial to derive pitch information from MIDI performance data, one is still facing problems in the domain of audio signal processing. However, monophonic pitch tracking and transcription into common practice notation (CPN) work well these days, at least for a reasonable number of instruments including the human voice. But the area of rhythm transcription without prior knowledge of the score is still considered to be a hard problem. We believe that solving this particular subset of Music Information Retrieval (MIR) would immediately enhance existing transcription systems, for example in commercial sequencers or in algorithmic composition and synthesis tools such as MaxMSP[3], OpenMusic (Agon, 1998) or Common Music (Taube, 1991). Our automated transcriber addresses the community of composers and musicians who use the above applications, but we hope that it will also appeal to musicologists and scholars in the fields of MIR, performance analysis and ethnomusicology, who might find it useful in support or as a replacement for the typical hand-editing and time-

[3]see the recent MaxScore extension by Didkovsky and Hajdu (2008)

consuming manual transcription process.

It poses also a challenge for composers themselves, to transcribe their vision into musical notation. Perhaps the reason for this lies in the limitations of Western CPN. Some of the musical phenomena are almost impossible to capture.

One main limit of CPN is the inability to communicate tempo modulations with the same level of precision that is achieved for notating pitch structures. But it is the modulation of a musical tempo, sometimes called *tempo rubato*, that is at the heart of every musical performance. If tempo modulations are central for music performances, then they must be central for composers as well, because such modulations would occur naturally in their minds as well. Within CPN, for example, *tempo rubato* can only be indicated in words; nothing else in the score will explicitly show performers how to bend the musical time in the characteristic manner. What are the musical reasons for performers to slow down or to speed up? What are the musical circumstances that lead to changes in tempo? What musical factors lead to the decision to start playing in a certain tempo? What is the extent of the influence of certain schools of playing that often emerge from the musical practice and aesthetic views of a certain period, such as the art of keyboard playing of Baroque music in contrast to that of the Classical and Romantic era? This is of course related to the development of the musical instrument itself but also to the development of a particular musical style. When analysing music that is not written down, can it be scientifically appropriate to make transcriptions? Would it be useful to transcribe music that is composed instantaneously, i.e., improvised music, which relies only on oral traditions. Or what about music that has a composed structure trying to circumvent the traditional problems of notation, e.g. space notation, aleatoric music. Even if the technical limits to music transcription can be overcome, which we would like to show in this thesis, there are certainly new aesthetic questions arising from the use of this new technology. Could our approach be useful within the current research on musical performances? Are transcriptions a useful source for the study of musical performances or are they perhaps dragging the attention away from important musical and perceptual aspects? How could expressive timing be best represented and what could be learned from the analysis of specific performances? Is the difference between the notation of rhythms in a score and their manifestation in a performance leading to answers about how we engage perceptually with expressive timing and how we perform and listen to music that unfolds in time?

We need to ask about the relation between a piece of music written as a modern Western score and its performance. How precisely do they differ? We might find an answer if we scrutinise the elements of music that are not directly expressed with CPN. Apart from the well-known primary properties of tones or musical entities as written in a score, such as pitch, duration, occurrence (position on the time-line), loudness and spectral content, like tone colour, formants, etc., there are also properties of their respective *contexts* as perceived by the composers, musicians and listeners. A score might only look like a collection of notes denoting their primary properties but what all performers strive to render is an aural image of the sounds, their connections and inter-relationships within their consciousness that transcend the bare bones of musical facts described in a score. It might seem as if a score denies the actuality of

time that music incorporates, in so far as music only exists during a performance. The score is not the music. Sounds may become music (Celibidache, 2008). Musicians have to create music by adjusting all relationships written in the score to every moment of the performance situation, for example a longer reverberation time in the concert hall leads the performer to choose a slower tempo in order to allow the music to emerge from that situation. A slight imbalance in dynamics within the ensemble leads the musician to adjust his loudness in order to reintegrate his voice into the ensemble. The question of choosing the right tempo is often debated within the musical community and tempo is dependent on all musical parameters, including instrumental qualities, or the harmonic complexity of a piece. If harmonic progressions are simple they do not exert much horizontal or vertical pressure, thus a faster tempo is possible, since the listener does not need much time to understand simple harmonic structures like for example I-V-I-IV-I-V-I . The rate of harmonic change, the harmonic rhythm, is another criterion that is related to harmonic pressure. If a piece is very chromatic with a fast harmonic rhythm, performers need to slow down the tempo, to allow them and the listeners to fully perceive the harmonic complexity. A high amount of harmonic pressure leads to a slower tempo. Tempo therefore is not only a "perceptual construct" (Cemgil, 2004), it is interwoven with all musical parameters of the score and more importantly the choice of tempo is related to context created by all parametric and time-dependent layers involved in a performance. The score and what can be learned from it, for example the harmonic complexity, give only hints. The music emerges from the moment of the performance with all variables connected to it: room acoustics, loudness of individual players, articulation, phrasing, the way the performers listen to each other, the hierarchy of Hauptstimme and Nebenstimme, main and secondary voices, melody and accompaniment, contrapuntal structure, bass-line, instrumentation and many sound properties of the individual instruments (a spectrum of a French Horn is entirely different from that of a Violin, and even every Violin itself is unique). A performance is the constant interaction between the knowledge of the performer, his or her actions and the sound that emerges at every now-point during the performance. The knowledge of the performer and listener involves the past and present musical experience as it unfolds and includes also expectations arising from the complex interplay of the past and present musical situation. When reading the score, even without instruments, a musical experience can be made, singing individual voices and listening to their combinations with the inner ear and creating an aural image of the sound. Following questions into the experience of time in our consciousness when listening to music one soon enters the realm of phenomenology where Edmund Husserl (1966) made important contributions that in the course of history influenced musicians like the Romanian-born conductor Sergiu Celibidache. According to the musical phenomenology, what constitutes musical experience in our consciousness has many time-dependent and intertwined layers. In the case of a single musical line, which is central to the creation of counterpoints, i.e., composing lines together, we observe not only the primary properties of individual tones but also secondary properties: All notes participating in a line also have properties that connect them to all other tones of the same line: a tone has also a linear function, a scale function, a harmonic function (even if it is only a single voice), and a dynamic function, see Thakar (1990), who studied conducting with

Sergiu Celibidache. By the term dynamic function Thakar means the experience of impulse and resolution that is inherent in musical lines. Tones participating in a line also participate in gathering musical energy, leading towards a climax, or they participate in the dissipation of that energy by contributing to the resolution. Thus, at any given time tones have a dynamic function. Although this phenomenon is central to music, as far as we know there has been no systematic study of its perceptual impact, let alone attempts to measure and quantify the dynamic function solely on the basis of a musical score. One reason might be that the dynamic function is, like the tempo, dependent on many musical factors. Tempo, intensity, thus individual phrasing and ensemble balance seem to be the most important components of the dynamic function when it comes to a live performance. The study of music recordings and field studies of live performances are required as well as advances in technology, like polyphonic voice tracking and extraction from audio signals, in order to have access to the primary properties of tones of individual players. Close microphone techniques in a studio environment might be the only currently feasible option for conducting a field study of the dynamic function. The linear, scale and harmonic functions of tones in a line can be derived from a knowledge of the score. We claim that these functions are important for the choice of tempo at any moment during the performance. A systematic study of their effects is needed in order to shed new light on the question what tempo really is: "...it is still unclear what precisely constitutes tempo and how it relates to the perception of the beat, rhythmical structures, pitch, style of music etc.. Tempo is a perceptual construct and cannot directly be measured in a performance." (Cemgil, 2004, pp.5). The phenomenologist would say that tempo is being born out of the necessity of realising the dynamic function following the build-up and release of musical energy that is taking place through a synthesis of all musical factors at any moment of a performance and where the past events influence the present actions and where the experience of the now-point leads to expectations and preparations for future musical events. Clearly these circumstances render every performance unique and unrepeatable.

This relativistic notion of musical tempo is related to the Buddhist philosophy of emptiness including the relativity of time, notably within the Sautrantika school. They argued for the "untenability of any notion of independently real past, present and future. They showed that time cannot be conceived as an intrinsically real entity existing independently of temporal phenomena but must be understood as a set of relations among temporal phenomena. Apart from the temporal phenomena upon which we construct the concept of time, there is no real time that is somehow the grand vessel in which things and events occur, an absolute that has an existence of its own." (Gyatso, 2005, p.60)

From the phenomenological and Buddhist angles, there is a danger of misunderstanding time and music through the practice of written-out scores in so far as a score might lead to the illusion of time as a "grand vessel" (empty staff lines) in which musical events (notes) are being placed as if an absolute time-line existed and thus the score and all its musical events would have an absolute and independent existence of their own. From this perspective it seems to be a great challenge and perhaps even impossible to reverse the process of a musical performance, i.e., starting with the performance data and inferring a score in CPN from them. But despite

these philosophical concerns we believe that we have found a viable and practical solution to the problem of performance transcription.

From this outline of (Western) musical practice specifically with regard to the dynamic function of sounds, it is evident that recordings[4] of musical performances show important changes in tempi happening on local time-scales that are bound, not exclusively, but for example by harmonic pressures, phrase structures and motive boundaries. Pianists are taught to virtually *sing* through their instrumental playing, instrumentalists breathe physically before the beginning of a new phrase in order to play *cantabile* or in order to pick-up the right tempo. Breathing helps the musician to phrase a piece naturally as if the instrument were his voice. Very often this support technique of vocal phrasing is audible on concert or even on studio recordings. Glenn Gould delivers an extreme example as he really sings along important musical lines. He also sometimes conducts himself with one hand while the other one is playing solo, so he demonstrates through gesture and voice the embodiment of musical expressive timing while playing on an instrument physically detached from the human voice. On an even lower time-scale there can be noise interfering with the musical time-flow originating from minute imprecisions due to the biological motor functions of musical players. The removal of such temporal noise originating in physical imperfection is what most musical training is addressing notably, through repetition at a slow tempo and by using various rhythmic techniques, e.g., playing a continuous 16th-note passage at a slow tempo with dotted notes[5]. On the other hand, timing deviations like *tempo rubato* are not written out explicitly in a musical score within the Western tradition. Instead musicians use their immediate musical instincts and the experience derived from culture, education and their own practice, to create situations where music can emerge. Of course, the score in all its minute detail *is not the music*. Nor can any other graphical or abstract representation

[4]During his lifetime Celibidache refused to release recordings of his orchestral performances on the grounds that the microphone is not capable of capturing all frequencies within the room, that it also modifies the sound spectrum due to its own impulse response, and most importantly because recordings are not listened to in the same room where they had been recorded. These deformations of the original sound render, from his point of view, all recordings into "bad photography". Celibidache commented, with typical wit: "Wouldn't you prefer to dance with Brigitte Bardot herself rather than with only a photograph of her?", see Fischer et al. (1986a), Fischer et al. (1986b) and also Celibidache (2008).

[5]Oswald (1997, pp.71) reports how Glenn Gould adopted a special training technique taught by his teacher Alberto Guerrero. He quotes William Aide, another piano student of Guerrero and a friend of Gould: "Finger-tapping is a lowly, obsessive, and cultish exercise for acquiring absolute evenness and ease in tricky passage work. It eliminates excess motion in the hand and ensures intimate tactile connection with the pattern in question. I will explain the practice in its simplest application. Take the notes D, E, F sharp, G, and A, for which the right hand fingering is thumb, 2, 3, 4, 5. The hand position is the natural one assumed when the arm and hand hang relaxed from the shoulder; the second knuckle is seen to be the highest point. Rest the finger pads on the key surfaces of the notes D, E, F sharp, G, and A. The left hand taps the fingers successively to the bottom of the keys. The right fingers are boneless; they reflex from the keybed and return to their original position on the surface of the keys. The left hand should tap near the tips of the right-hand fingers, either on the fingernails or a the first joint. The motion of the tapping should be as fast as possible. The second stage of this regimen is to play the notes with a quick staccato motion, one finger at a time, from the surface of the key, quick to the surface of the keybed, and back to the surface of the key. This is slow practice, each note being separated by about two seconds of silence." Oswald then continues: "Guerrero claimed to have hit on the finger-tapping method independently after attending a circus where he saw a three-year-old Chinese boy do an astounding dance full of breath-taking intricacies. Guerrero went backstage to meet the child and asked his trainer for the secret. The teacher-trainer demonstrated how he placed his hands on the child and moved his limbs, while the child remained still and relaxed. Then the child was asked to repeat the movements by himself. [...] Ray Dudley, another piano student, [...] testifies that Gould finger-tapped every Goldberg-Variation before he recorded it ... Gould boasted to Dudley that tapping the complete Goldberg-Variations took him thirty-two hours."

of recorded audio data by definition reconstitute what music in its essence really is. The reason for that is philosophical, because all representation can only be analytical. Only the music itself, experienced in the here and now, *and only if* the right conditions for her emergence are set by composers and musicians alike, can be transcendental in a metaphysical sense.

We therefore need to understand more fully all aspects of expressive timing in musical performances and how they emerge from a musical piece encoded as a score. We believe that the development of an automated score transcriber based upon onset data alone and without previous knowledge of the score can serve as a tool for research in music cognition and performance practices. With such a tool it will be possible to scrutinise the *musical reasons* for particular performers' decisions in the domain of expressive timing. Patterns and individual habits of players and perhaps also schools of interpretation can be traced from applying an automated transcriber. One could argue in this case for the use of a score follower but that would first require a manual input of the score data and secondly, existing score followers typically require a considerable amount of hand-editing during the performance in case a musical event has been missed[6]. A transcriber without score knowledge also has the advantage of being applicable to non-notated forms of musical practice, e.g., improvisation, although aesthetic questions remain to be discussed, e.g., the notion of choosing a metre that implies certain accentuations of musical events. 20th century avant-garde composers have very often written a note into the score saying that the bar-lines are merely for synchronisation purposes and under no circumstances mean an accent on the first beat of a bar (e.g., Ligeti, 2nd String-Quartett, *Lux Aeterna* et al.). This means that CPN is not a fixed set of rules implying specific musical practices but rather that over the centuries a graphical system has evolved which is capable of capturing complex musical structures of very different styles and that this graphical system is like any other system or *textual framework* as such open for re-interpretations, especially if certain aesthetics demand a change of common practice. Finally, we hope that the tool we describe in this thesis can be of use in order to study the dynamic function of sounds and their influence on expressive timing in performances.

Automated music transcription often divides its task into two main problems. One is tempo or beat tracking, where the notion of tempo is reduced to a mechanical one, and tempo is defined by beats-per-minute. Note how this is in contrast with the phenomenological view of our previous discussion. The other problem is the quantisation of onsets given a tempo derived from a tempo tracking process. It is often stated in the literature that this causes a "chicken-and-egg" situation: a successful quantisation relies on a known tempo and a tempo tracking algorithm relies on outcomes of the quantisation process, see Cemgil (2004) and Raphael (2001). We need to see if the methodological divide into tempo tracking and quantisation is absolutely necessary. However, there are obvious areas in score recognition where even the latest methods of transcription are not applicable. We mean in particular certain polyphonic textures in pieces by Ligeti (*Atmosphères*, *Requiem*) or Xenakis (*Metastaseis*, *Phitoprakta*) where individual voices are merging into great masses or clouds of sound. Clearly, where the listening experience is confronted with non-deterministic musical textures it will only be able to give statistical

[6]Arshia Cont presented at the ICMC 2008 a new score follower that promises to be more robust

descriptions of the sounds. No transcription system is known yet or even conceivable that would be able to find the original CPN score of a micro-polyphonic passage of Ligeti.

1.3 Previous Publications

One of the main contributions of this thesis is the use of the Farey Sequence in connection with various filtering algorithms as a general model for musical rhythm. The method is specifically aimed at those forms and practices of music that are based on an underlying pulse. This is generally the case in music performances that are based on the concepts of bar and metre, but the method applies also to non-Western music traditions such as African polyrhythmic music, which is totally unrelated to Western music theory. This hypothesis was first published by Boenn (2007b). Additional information about this subject and the related music theory will be covered in chapter 2. In order to find a common ground for modelling a large amount of different styles of music and in order to represent even the most intricate forms of polyrhythms we have looked for an appropriate mathematical representation that would enable us to compute such rhythmic structures both in terms of analysing real world performances but also in terms of offering composers and musicians a comprehensive and flexible tool for the creation of rhythms in general. The use of the Farey Sequence as a model for musical rhythms in different styles has been initially discussed (Boenn, 2007b) and an object-oriented programming framework for composing rhythms with the Farey Sequence was presented. Using this framework we had created a few examples from music history that were analysed and represented by means of *filtered* Farey Sequences. Sections 3.2 and 3.3 are elaborated presentations of the methods already outlined in this paper (Boenn, 2007b), together with a new focus on the various filtering processes that can be applied to Farey Sequences in order to model musical rhythms. In our thesis, the method is also expanded to include the modelling of musical metre. In addition, the author recently wrote a suite of Csound opcodes[7] (Boulanger, 2000) that facilitate the modelling of rhythms by means of Farey Sequences.

In audio editing software, the process of quantising music performance data is usually understood as a procedure for mapping the onset times of note events to an equidistant grid representing a simple sequence of isochronous note values, for example semiquavers. This oversimplification of the musical performance data leads to unwanted results, such as the false overlap of note events that were originally separate. More importantly, it is very hard if not impossible for existing algorithms to take the musically expressive timing deviations of performers into account. Our efforts concentrated on finding a solution to this problem and to develop a quantiser that does take into account expressive timing. We have already published first test results of our own quantiser that were based on an earlier version of the quantisation algorithm (Boenn, 2007a), and using two different recordings by Glenn Gould of Johann Sebastian Bach's *Aria* of the *Goldberg Variations*, BWV 988. Now we are able to present a refined and expanded version of the algorithm in chapters 5 and 6 of this thesis. In addition, new experimental results are presented and discussed. The main restriction of the previous method

[7] The following Opcodes are now part of Csound, version 5.13: *GENfarey, tablefilter, fareylen and tableshuffle*

of quantisation (Boenn, 2007a) was that the information of the downbeat had to be given in advance. With refined and new methods presented here we can now show possible ways to lift this restriction, which have been found via experimentation. These strategies are based on the new development of an onset segmentation process and on a new process to concatenate the optimum outcomes of the quantisation into a continuous metrical timing representation of a musical performance.

The notion of retentional rhythms and their use for composition has been previously published (Boenn, 2008). The thesis develops this idea further in chapter 7, where we also demonstrate the relationship of retentional rhythms with the Farey Sequence.

We will also add discussions of our findings in relationship to recent publications in the area of auditory perception (Large, 2008; Large et al., 2010; London, 2004).

1.4 Contributions

In this thesis we are able to show that the Farey Sequence can serve as a general model for musical rhythm and metre. We found that the Farey Sequence is similar to the structure of metrical subdivisions of beats, bars and even higher-level elements of musical form. We thought it would be interesting to find out if the Farey Sequence could explain the formal structure and rhythmic building blocks of entire sections of musical pieces, for which we will give evidence. Furthermore, we will show how the Farey Sequence is helpful in order to analyse the rhythmic micro-structure of human performances of music. Because many references will be made to the Farey Sequence, we would like to give the following definition first (see Hardy and Wright, 1938, p.23):

A Farey Sequence F_n of order n is a list of fractions in their lowest terms between 0 and 1 and in ascending order. Their denominators do not exceed n. This means, a fraction a/b belongs to F_n, if

$$0 <= a <= b <= n \tag{1.1}$$

with $a, b, n \in \mathbb{N}$. In F_n, the numerator and denominator of each fraction is always coprime, and 0 and 1 are included in F_n as the fractions $0/1$ and $1/1$. For example:

$$F_5 = \{\frac{0}{1}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1}\} \tag{1.2}$$

Note that there is an infinite number of Farey Sequences and that the following relationship holds:

$$F_n \subseteq F_{n+i}, \tag{1.3}$$

with $i \in \mathbb{N}$.

We will demonstrate that the Farey Sequence can be used as an underlying grid structure for the analysis of musical performances. Individual elements of the Farey Sequence can serve as discrete time points to which the musical onset data of performances can be mapped. The Farey Sequence makes it then easy to translate the quantised time points to the specific val-

ues for musical notation (CPN). This process is generally known as quantisation and we will compare our output to the recorded musical performance and to a human transcription of that particular performance. The Farey Sequence is also useful for modelling other aspects of musical rhythm and metre, for example the so-called metrical tempo grid, which has been proposed by London (2004). A metrical tempo grid shows how specific musical tempi are correlated with psychological timing thresholds. It can reveal how these timing thresholds guide and restrict the selection of metres and rhythmic values for the composition and notation of musical rhythms in Western music.

From the composer's point of view, the Farey Sequence is very interesting because it lets one model a diverse set of the rhythmic components of musical styles, for example Renaissance counterpoint and African Polyrhythm. Farey Sequences can translate directly into musical rhythms by letting their elements determine the onset times of musical events. It is then interesting to explore the different rhythmic structures that emerge from Farey Sequences. For example, for different values of n, F_n exhibits different rhythmic properties. F_n can serve as an encoding for musical polyrhythms and in general as a representation for any conceivable compound rhythm that is built over a grid of simple pulsation, or which is embedded in a hierarchical metric structure that is commonly employed in Western and non-Western music (Arom, 1991; London, 2004). Various filters for Farey Sequences are being discussed, which open up many possibilities for the modelling of rhythms and metrical structures.

We will also see how Farey Sequences can relate to memory representations of musical percepts. This is the basis of our second contribution. In particular, we will introduce the concept of *Retentional Rhythms* that is based on the philosophical ideas of Edmund Husserl. The term is directly related to his phenomenological analysis of inner time perception with continuous series of retentions and protentions originating from the now-point. When applied to the phenomenon of music listening, such series of retentions contain rhythmic timing information that can be correlated with each other. As a consequence, important aspects of the perception of musical rhythms will emerge. In particular, there are phenomena of resonance occurring in retentional maps of perceived rhythms, which can be used to explain certain aspects of rhythmic accentuation, like, for example, metrical downbeats, up-beats and syncopated structures[8]. The term resonance is understood as the combined peak amplitudes of at least two oscillations who are in phase at certain points in time and whose periods are equal to inter-onset intervals (IOIs), which are derived from the perceived rhythmic timing information. It will be interesting to see how these outcomes can be related to current research of models of neural oscillations. There are resonance phenomena occurring in these models that can explain important aspects of the perception of rhythmic stimuli (Large, 2008; Large et al., 2010).

[8]There is a debate amongst musicologists about the terms stress and accent. The main question is how the two are different from each other. For clarity we will only employ the term accent.

1.5 Outline of the Thesis

The content of our thesis is structured in the following way. In chapter 2 we present the background and related work for our research into automated quantisation and transcription of onset data from musical performances. Chapter 3 explains the Farey Sequence as a model for musical rhythm and metre. We look in detail at filtered Farey Sequences and their role within the analysis of musical compound rhythms by using a large number of examples from various musical styles. The following chapter 4 describes our experimental framework, what we are going to test, how we arrive at our results and why we think they can provide evidence for our main hypothesis. Chapters 5 and 6 are discussing the core algorithms of our work, namely the grouping of onset data into duration classes, and their subsequent quantisation and transcription into Western score notation. We provide experimental evidence by using different styles of music and present complete transcriptions of Glenn Gould's recording of the *Aria* of Bach's *Goldberg Variations* at the end of chapter 6. The following chapter 7 discusses a generative method of composing and analysing musical rhythms on the basis of the phenomenology of inner time-consciousness by Edmund Husserl (1966). We conclude our thesis with a description of future work in chapter 8 followed by a final summary and conclusion in chapter 9.

Chapter 2

Background and Related Work

2.1 Introduction

In order to define the requirements for a computer model of musical rhythm it is useful to understand how rhythms are being represented and communicated in written form. Various symbolic languages have been developed over history, but one of the most common forms in use nowadays is Western music notation. This is also the form of notation normally used in musicology for the transcription of rhythms of real musical performances. A brief survey of other important representations of rhythm and metre will also be given.

For the collection of test data used by our quantisation and transcription algorithms we need to understand and apply current techniques of onset detection and onset extraction from various data sources. We also would like to know how much noise is likely to occur in the data when those methods are being used.

Important aspects of temporal perception of musical phenomena are taken into account for the design of our model. We will also analyse the relation between a score written in CPN and its musical performance, which always involves expressive timing, an area also called *agogics*. We will briefly explain this aspect of musical performance, along with a close look at ornamentation. Musical ornaments give performers many opportunities to use expressive timing. Therefore we have chosen a richly ornamented piece from the repertoire of Baroque keyboard music, the *Aria* of the *Goldberg Variations*, in order study various challenges for music quantisation and transcription.

For the analysis of musically performed durations we found that the Gestalt theory offers viable approaches. One key element of our model uses an original method of grouping and classification of durations, which is related to k-means clustering.

Various results from the research into musical metre and rhythm analysis and perception are being presented. Finally, we will be looking at the current state-of-the-art of quantisation and tempo-tracking.

2.2 Representations of Musical Rhythm

In this section we will present various symbolic or graphical forms of representing rhythms and metre. Because our end-result of rhythm transcription shall use the Western common practice notation (CPN), and in order to analyse the relationship between written music and its performance, we will begin with the perhaps most precise and versatile form of symbolic notation.

2.2.1 Notation of Rhythm and Metre

In CPN rhythms are represented by a set of ratios with small integers, which encode note durations in relationship to a larger reference unit, see Chew and Rastall (2001). For example, a crotchet lasts one fourth of the time of a semibreve. But this reference duration has not the unit of physical time, because event durations in CPN are relative and not prescribed using explicit physical time values. A musical tempo is therefore not part of the information encoded in CPN. However, it is possible to capture the temporal relationships between an isochronous pulsation, for example a musical beat, and the actual note on- and offsets. Herein lies the function of bar and metre. In music, pulses and beats have the special property of being extended in time. A beat, therefore, has a relative duration just like an ordinary note has. One of the main conventions of CPN is to align note durations with a hierarchic metrical grid, i.e., a grid composed of durations. Pulsation and metre are sometimes referred to as the simplest form of rhythm (Arom, 1991). Metres are in fact indicated by time signatures, which are also integer ratios relating to the semibreve as a reference. The purpose of a time signature is to define the duration of a bar and how this bar is composed out of smaller units of durations. The basic metrical grid of beats is formed within a bar. The first beat in a bar is called the *downbeat*. Later we will see how a metrical grid can also span across groups of bars. The following figure 2-1 illustrates how note durations of a musical piece relate to a metrical grid that is defined by the time signature *and* by the compound rhythmic surface of the music. Heavy downbeats, weaker beats and further subdivisions of the beat are represented by particular rows of dots below the piano staves. The interplay between strong and weaker metrical positions in time is a feature of metrical hierarchy. The musical structure can contradict this regular scheme by placing accents on weaker parts of the metre, a process known as syncopation. A single dot denotes onsets of a particular metrical level – a representation that is widely used in music theory (Lerdahl and Jackendoff, 1996; London, 2004). Although a beat carries a duration in the score, there is no direct inclusion of the actual tempo nor of absolute physical time. This has not changed even with the advent of the metronome in the 19th century. Although from then on composers could include absolute timing information in their scores, the metronome value of the beat was only given at the beginning of pieces or sections as an *indication* of the tempo in a unit called beats per minute (BPM).

The following figures 2-2, 2-3 and 2-4 show the basic set of symbols used in CPN for note durations, rests and time signatures. We also give the name of the symbol and the integer ratio that represents the note duration in relation to a reference duration. In CPN, this reference

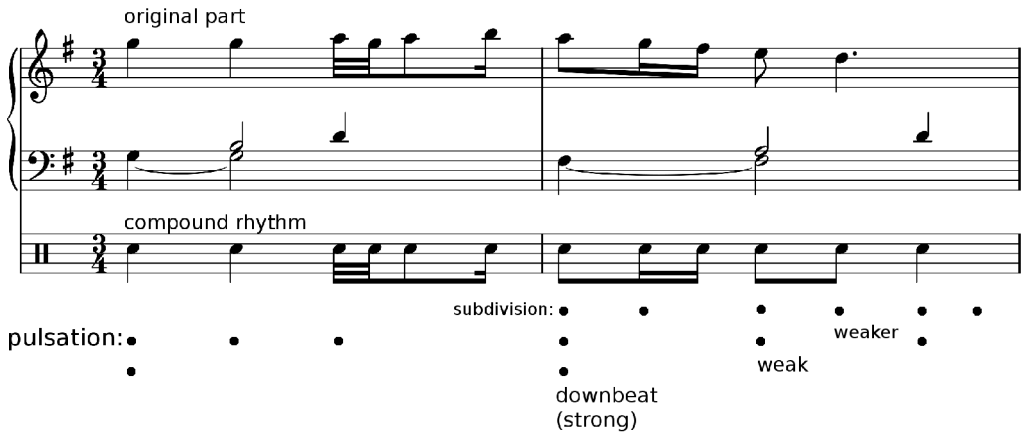


Figure 2-1: Beginning of the *Aria* of J.S. Bach's *Goldberg Variations*. Rows of equidistant dots denote pulsations on a specific level within the metrical hierarchy introduced by the nature of the 3/4 metre. The upper row represents 1/8 note pulsation, the centre row represents the beat level in 1/4 notes, and the lowest row indicates the strong downbeats, which form a pulsation in dotted 1/2 notes. Notes are found on weaker metrical levels if they do not coincide exactly with the downbeat.

duration is always the semibreve, 1/1. A semibreve comprises four crotchets (1/4), eight quavers (1/8), and so forth. Rhythms can be written down using a combination of the duration symbols. They are always aligned to the underlying pulsation or metrical grid.

There are multiple ways of expanding the basic set of simple integer ratios given above. First, by using slurs between notes it is possible to concatenate them and to express compound note durations. As a convention, one must do so always when a note is so long that its duration crosses a barline. A dot placed after the note head increases the duration of the note by 50%. Two dots after the head increase the duration by 75%. The use of dots rather than slurs is a shorthand that prevents the notation from being too difficult to read since the dot extension of a duration is used very often. Another possibility of expanding the basic set of durations is to write so-called *n-tuplets*. These are divisions by n of the semibreve where n is not equal to a power of two. The ternary subdivision of the beat is as important as the binary one; so are nested subdivisions of the reference unit in twos and threes. Although it would be sufficient to express all kinds of n-divisions as ratios with reference to the semibreve, for example three 1/12 notes for a quaver triplet, CPN rather uses specific shorthands for *n-tuplets*. Figure 2-3, second staff from below, shows an example of groups of triplets (12-tuplet) in a 4/4 bar. The number 3 below the beam indicates: "Play three eighth notes during the time-span of two". This is a shorthand for the rhythmic proportion 3/2 and it is applicable to all basic note durations. Nested tuplets can also be used as illustrated with a 9-tuplet in figure 2-3, 6th staff from the top. The total duration of the parent tuplet is also aligned to the underlying metrical grid. In most cases, all tuplets are aligned to the beat level. According to this rule, one is not allowed to shorten a tuplet, for example using only 2 of 3 notes of a triplet, or to use only one single element of a tuplet on its own.

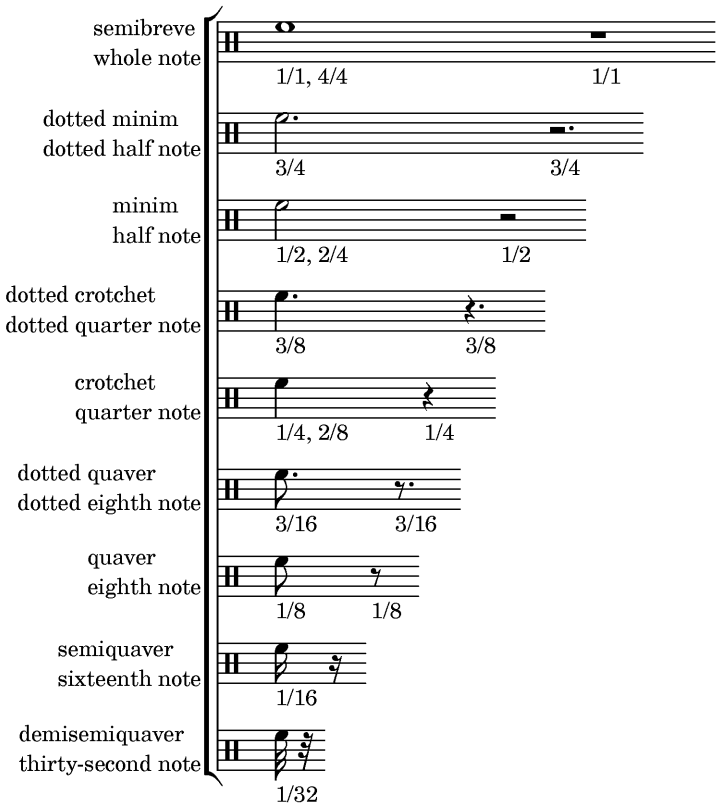


Figure 2-2: Notation of note durations and rests as ratios in relation to the semibreve (whole note) as a reference but without reference to absolute physical time or musical tempo.

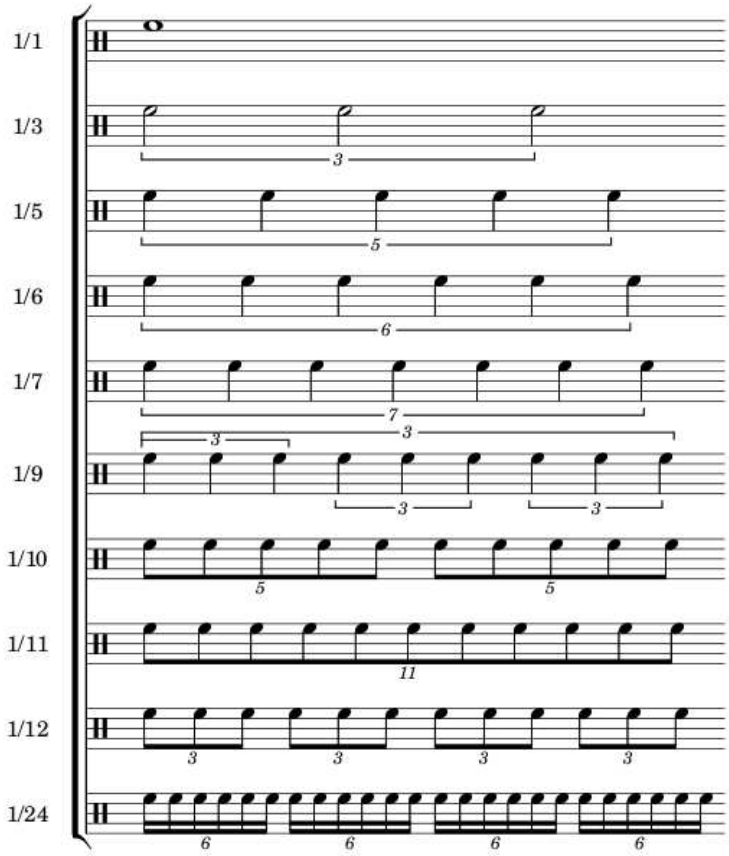


Figure 2-3: Principle of subdivision in CPN as extensions of basic durations given in figure 2-2. Higher prime number subdivisons are not often used in Western music, whereas subdivisions based on prime factors of 2 and 3 are very common.

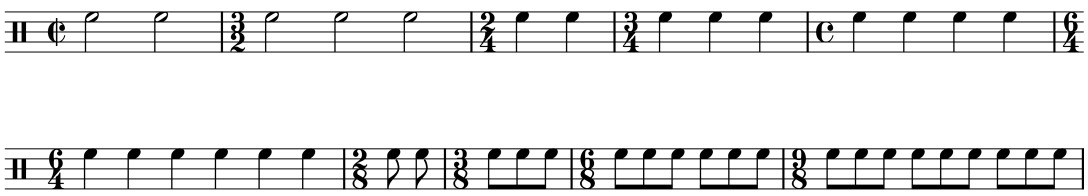


Figure 2-4: Common time signatures for various metres in CPN. The notes display the pulsations on the beat level.

CPN uses fairly simple ratios in order to represent musical durations and rhythmic structures. It was also found in experiments that humans categorise groups of durations according to their vicinity to simple ratios and maps of these categories can be found for example in Desain and Honing (2003). Western music theory often discusses rhythmic composition techniques on the basis of the ratios that are possible to notate within CPN, for example Stockhausen (1988). In his article, Stockhausen draws parallels between rhythmic textures, tempo relations and pitch intervals of the overtone series in order to build a unified serialistic approach to pitch and rhythm composition.

2.2.2 The Piano-Roll Notation

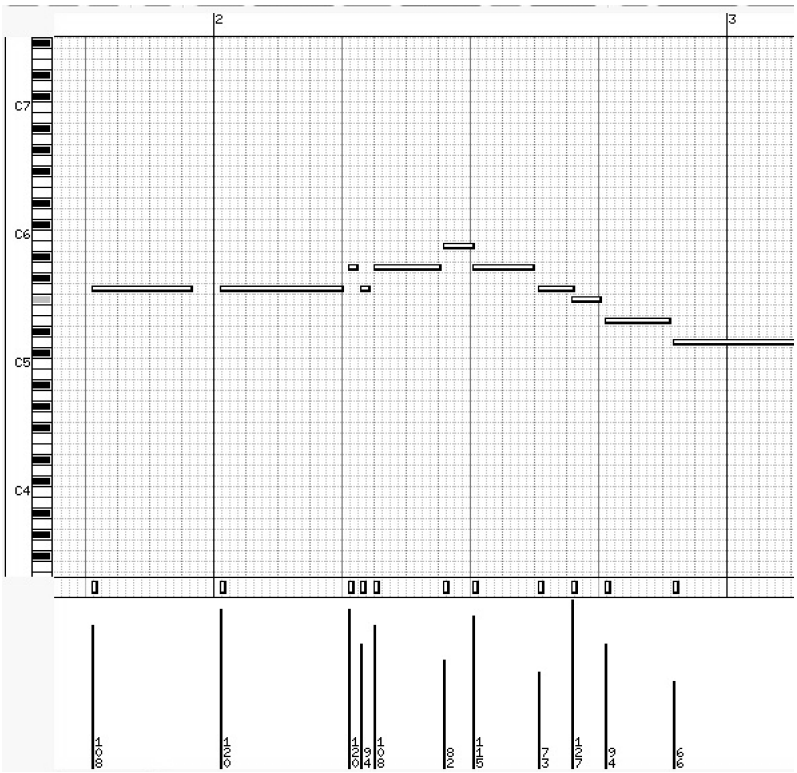


Figure 2-5: A MIDI track in piano-roll notation. The view at the bottom indicates note velocities. They are equivalent to the force excerted on the key by the player

This is a common notation practice in MIDI sequencer software. Notes played on a MIDI keyboard translate into note-on and note-off events according to the action of the player on a particular key, for instance the velocity of the key's movement is measured and sent together with the MIDI key number and the MIDI channel number as a single note-on event. The delta time between the note-on and note-off events is measured and the timing, key and velocity is drawn on a score view that resembles a piano-roll, see figure 2-5. In this view the note duration is represented by a little rectangle and its width in pixels scales according to the length of the note.

The practice to represent note durations by using one dimension in space is called *space notation* and was perhaps first used by John Cage in his hand-written scores of piano music. In the *Music of Changes* for example, 1 inch on paper is equal to 1 second of the performance. In the MIDI piano-roll notation, the vertical position of the note rectangle indicates the pitch of the note. Typically, MIDI pitch numbers map to the equal-tempered chromatic scale of Western music. More advanced programs allow the user to change this mapping and to employ a different tuning system, for example by using a scale in just intonation (Partch, 1979). In a MIDI sequencer, the tempo of the performance is assumed to remain constant over the background of inaudible metronome ticks and all note events are drawn in proportion to this underlying fixed tempo. With regard to musical tempo, the sequencer's representation is completely independent from any changes the musician might introduce when playing musically and expressive. One can only employ a so-called tempo track in order to hand-edit tempo changes within the MIDI sequence. But, this is only possible after the performance has been recorded. The MIDI tempo events on the separate tempo track will affect only the way in which the delta time between all other MIDI events is interpreted by the sequencer program.

2.2.3 Necklace Notation of Rhythm and Metre

London (2004) and Sethares (2007) take advantage of necklace notation for musical rhythms and metric hierarchies. This form of notation dates back to 13th century Arab music theory, when Safī al-Dīn al-Urmawī published it for the first time in his *Kitāb al-adwār* ('Book of Cycles') (al Urmawī, 1980; Sethares, 2007). The main idea is that musical metre is a cyclical repetition of patterns of attentional energy focused by the listener on specific moments within a musical metre (London, 2004). A cyclical representation supports the notion of repetitive patterns. Although music works with the principle of variation, any variation of rhythmic patterns are carried out on the backdrop of culturally accepted and individually learned metrical patterns. London (2004) calls this process *entrainment* of metrical patterns by the listener. Small changes of musical parameters will keep the listeners attention while the inherent stability of the underlying metrical hierarchy helps the listener to fully appreciate the changes and to grasp their aesthetic meaning. The following figures depict such stable metric patterns from different musical cultures. A circle represents a note onset event as well as an expectancy peak in an entrained pattern of musical metre. In necklace notation, time flows continuously and clockwise around the circle. Figure 2-6 shows the Arab rhythm *thāqil thāni*. Large dots indicate the initial time or foot that is always sounded, medium dots denote medial time units, optionally sounded, whereas the smallest dots indicate the final time unit that is always silent (Right, 2001). Figure 2-7 shows the Ghanaian *Ewe* rhythm. Variations of this rhythm appear in other African cultures as well, for example in Central Africa and in Nigeria. The difference between them is that their particular version starts at a different time point within an otherwise identical rhythmic pattern (Sethares, 2007).

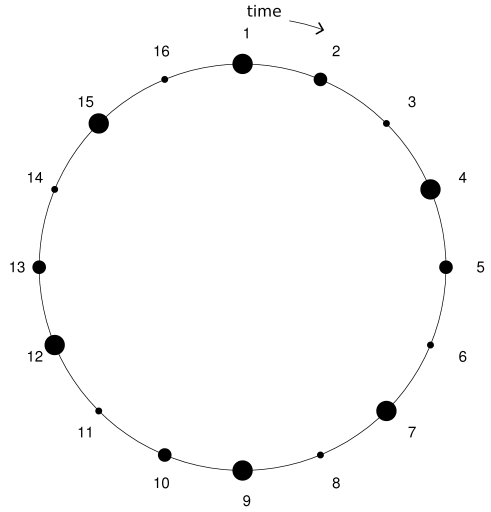


Figure 2-6: The Arab rhythm *thāqil thāni*, after Right (2001).

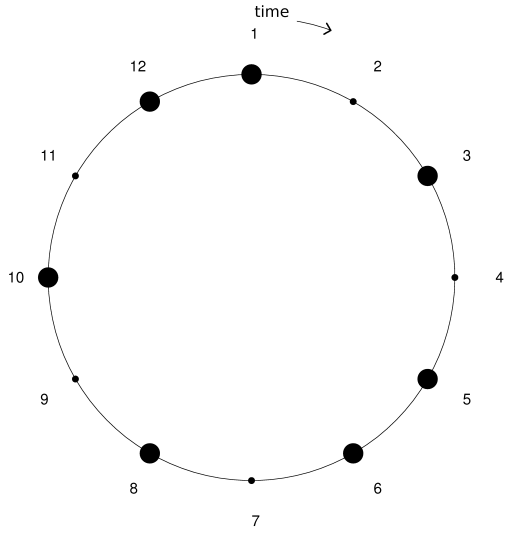


Figure 2-7: The *Ewe* rhythm from Ghana, Africa, after Sethares (2007).

2.2.4 Adjacent Interval Spectrum

The analysis of performed musical durations is one of the key tasks of our proposed method of quantisation, which allows to capture subtle tempo changes and expressive timing in human performances. To achieve this, it is very helpful to visualise musical performance data by using an adjacent interval spectrum, an original method developed by Kjell Gustafson (Toussaint, 2004). The main idea is to plot timed durations as boxes against the vertical axis, and the sequential order of their occurrence against the horizontal axis, like for example in figure 2-8. When looking at the interval spectrum, it easy to compare the plotted durations with each other and to detect quasi equal durations, which, despite local variations, indeed belong to the same kind of duration in CPN. In figure 2-8 for example, one can see that the durations number 4, 6, 9 and 10 have values close to 0.08. There is also a significant distance between them and other durations. But, over the course of these two bars they seem to become slightly longer each time. A comparison with the score in figure 2-1 can then reveal how a particular musical event has been timed by Gould and how it was put into a musical context, for example a tendency to slow down with the phrase ending at the end of bar 2 is reflected across various duration classes. And one can see in figure 2-8 that this slowing down starts already at the beginning of bar 2 with duration number 6. The last duration 11 seems to stand out from the others but it is according to the score the same note value, a crotchet, as the durations 1 and 2 from the beginning of the piece. This manner to deliberately bend musical time as a means of human expression by the performer is a wide-spread practice across Western musical culture and can also be found in other parts of the world. And although musical tempo seems to fluctuate all the time, a human listener has obviously no problems to follow these tempo changes, for example to tap the foot along, and take a delight in the bending of time as an important part of the musical experience. It is then for the design of an automated quantiser and transcription system an equally important but also challenging task to be able to follow these tempo changes accurately.

2.3 Onset Detection

The detection of note onset times in musical performances is crucial for the creation of experimental data sets. Previous research about tempo tracking and automated transcription used onset data from various data sources, like manual tapping, audio and MIDI recordings. They also used different musical styles. We expect that the detection and extraction of onset data will introduce noise. The question is whether the noise can be ignored for the purpose of rhythm quantisation. For it to be ignored we need to know more about the kind of errors and their approximate levels that are going to be introduced by the various methods of measurement.

2.3.1 Manual Tapping

When using a computer keyboard for manually tapping along an audio signal, Wright et al. (2004) reported latency and sometimes significant jitter from the keyboard. The best results

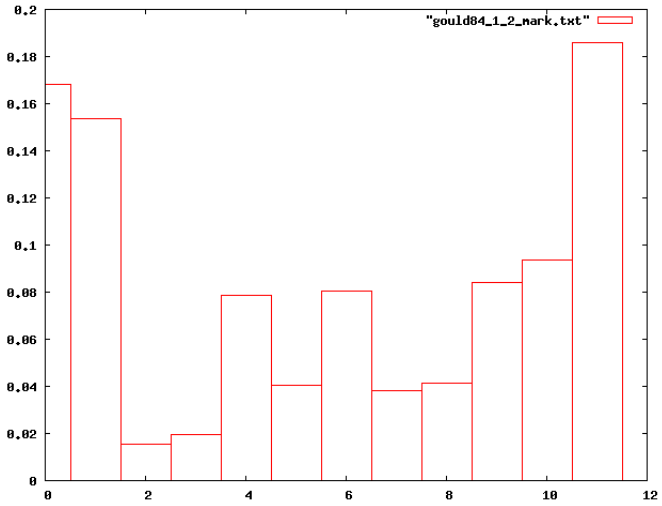


Figure 2-8: Adjacent interval spectrum of the compound rhythm of the first two bars of J.S. Bach's *Aria* performed by Glenn Gould in 1981, with boxes plotting the normalised durations of inter-onset intervals on the y-axis. Figure 2-1 shows the score.

of low latencies and jitter were achieved when the ALSA and JACK audio drivers had been set at low latencies on a Linux system configured for real-time audio. MaxMSP under MacOS X achieves comparable results when the "overdrive" option is turned off. Under those optimised conditions one has to take on average 21-22 ms of latency from computer keyboards into account. A quantisation algorithm has to be robust enough in the presence of jitter and should run preferably under a system configured properly for real-time audio. 22 ms latency is equivalent to the point-of-view of a musician controlling a sound source that is 7.55 metres away. The maximum rate for repetitive musical timing gestures, e.g., trills, is on average 14 Hz, which means that their components are approx. 71 ms long (Brown and Smaragdis, 2004). The fastest periods on average for purely rhythmic events is around 100 ms. Due to receiving acoustic feedback of their action we expect some influence of the latency on the player's performance at very fast tempi. However, players using the manual tapping apparatus can adapt in a similar way compared to their adaptation of large room acoustics. We do not expect any influence on players at moderate to fast musical tempi. This problem of latency disappears with the cancellation of acoustic feedback within the tapping procedure. Hearing only the tapping sounds from the computer keyboard itself might be enough information for a player to perform rhythms in combination with the tactile feedback. The person could listen via headphones to the original music he or she is tapping along on a computer keyboard. In this way the rhythmic information of an audio recording can be translated into onset times ready for further analysis, quantisation and transcription. For our experiments we used both approaches for onset recording: Tapping with acoustic feedback from a software synthesiser, and silent tapping while listening to rhythms presented via headphones. In both cases we used Csound on a Linux laptop.

The *times* Opcode in Csound

When timing events in Csound is required, for example external control data are received that trigger events in Csound, and we wish to record the exact time of the incoming event, then a sample-accurate resolution can be achieved by using the *times* opcode.

The Csound Instrument in the Appendix A on page 183 shows the application of the *times* opcode so that users could record rhythm onsets via computer keyboard. This recording can then easily feed into our quantisation and transcription programs.

2.3.2 MIDI

MIDI Files

According to The MIDI Manufacturers Association (1996), MIDI messages such as 'note on' and 'note off' in MIDI files are preceded by delta-time information encoded as an integer based on the number of ticks per crotchet or PPQ (pulse per quarter note). This PPQ parameter sets the MIDI timing resolution. The absolute length of a single pulse p is defined initially by the first tempo message of the MIDI sequence in microseconds divided by PPQ:

$$p[\mu s] = tpo[\mu s]/ppq \tag{2.1}$$

The tempo message is encoded as the length of a crotchet in microseconds. In MIDI file format 1 the tempo messages are located on a so-called tempo-track. If no tempo message is present, then the default tempo value of $5 \times 10^5 \mu s$ (= 120 BPM) is assumed. Since the PPQ value sets the timing resolution of a MIDI recording, quantisation already takes place by mapping the incoming note message to the last grid position with timestamps converted to a value in $[\mu s]$ divided by the pulse length $p[\mu s]$. Due to rounding errors that might occur, one must avoid low PPQ values combined with a slow tempo setting. The PPQ value together with the tempo message(s) of the MIDI file as the duration of the crotchet in μs allows a program to calculate the IOI or note duration between note-on and note-off[1] messages given as onset or offset information in $[ppq]$.

$$duration_i[sec] = (offset_i - onset_i)\frac{p[\mu s]}{10^6} \tag{2.2}$$

with i = index of the tuple of note-on and note-off onset times delivered by the MIDI stream. The onset times of a particular note within the sequence can be calculated cumulatively from the delta times of the previous MIDI events. This is necessary because the serial format of MIDI might place other events in-between a note-on message and its corresponding note-off event. Apart from PPQ, sometimes a SMPTE time-code in frames-per-second is used.[2] In order to trigger a change of tempo in a MIDI file so-called meta-events are being used. MIDI files of type 1 contain a tempo-track on which the tempo messages are recorded. This

[1] where note-offs are often encoded as note-ons with zero velocity

[2] `http://developer.apple.com/documentation/Java/Reference/1.5.0/doc/api/javax/sound/midi/MidiFileFormat.html` [Accessed June 2008].

ensures that the tempo of a performance can be changed independently from the delta-times recorded from the performance. The time-signature is encoded as a combination of its start-time, beats-per-measure (numerator of the time-signature), denominator of the time-signature encoded as the negative exponent of two, e.g., 2 = crotchet, 3 = quaver, etc., and number-of-midi-sync-messages-per-crotchet [default:24] [3] and number-of-demisemiquavers-per-crotchet [default:8]. If the quantisation option of a sequencer is switched on, MIDI file streams encode rhythms by using a hierarchical bar representation with beats and OxF8 sync-messages. Note that for subdivisions of the crotchet where higher prime numbers are involved the resulting rounding-off errors lead to timing errors of the encoded rhythm. Therefore, if higher prime-number subdivisions of the crotchet are needed for the composition, quantisation on commercial sequencers can pose problems if the PPQ value is not adapted to the subdivision of the crotchet. The PPQ value always has to be a k-smooth number (Berndt, 1994; Blecksmith et al., 1998) with k equal to the highest prime number subdivision of the crotchet. A positive integer is k-smooth if its prime divisors are $\leq k$. Of course, the PPQ and tempo settings of the crotchet have a relative effect also on time signatures with a 16th, 8th, or minim as denominator.

According to The MIDI Manufacturers Association (1996), in situations where a milliseconds resolution of the delta time is wanted or a precise correlation between MIDI events and video frames according to SMPTE time-code, then a different definition of the PPQ value is possible given by the *division* word of the MIDI-file header. If its MSB is set to one, then the following bits 14 to 8 represent one of the four SMPTE frame rates: -24, -25, -29 (30 with drop frame) and -30 in two's complement form. The second byte of the division word encodes the number of pulses per video frame. Therefore, with SMPTE frame-rate of 25 fps and 40 pulses per frame, all delta times would be in milliseconds. It is not advisable to use a lower timing resolution in MIDI files. For experiments in rhythm quantisation, a delta time resolution of 1 millisecond is sufficient. This is already the best rate MIDI can offer. A 1 - 2 millisecond delta time between events would be audible given laboratory based test outcomes, see Wright and Brandt (2001). Because MIDI uses a *serial* data transmission protocol, and because the maximum rate for messages is indeed approx. 1000 per second, there is a danger that musical chords accumulate latency between the first and the last note of the chord. An orchestral chord with 21 voices for example would carry a delay of 20 milliseconds between the first and the last note, but only if the full bandwidth of the transmission is used for the chord alone. Any further MIDI messages intertwined with those of the note events could easily increase that delay[4]. We will also take into account a certain haptic timing limit (Verplank et al., 2000), because they help to detect chords in MIDI and to distinguish them from fast gestures like trills and arpeggios, see also Kilian (2004). The IOI for the fastest trills played by trained musicians is around 71 milliseconds (Brown and Smaragdis, 2004). We will therefore filter those gaps between MIDI events, which are below 71 milliseconds. With this method we can determine, which MIDI

[3] this means that a status byte 0xF8 is sent 24 times per crotchet (System-Realtime-Message). Any higher PPQ value is achieved internally by subdivision of the inter-sync-message-interval. For a PPQ of 960, the sync-duration needs to be divided by 40.

[4] Modern MIDI data transmission via USB has overcome this problem. Also, internal audio synthesis based on MIDI files, for example with Csound, does not introduce latency caused by MIDI messages.

events are supposed to sound together, for example chords and fast arpeggios, but also two or more polyphonic voices who can, at any particular moment, share the same note onset time in the score. It will also deal with the problem of timing delays caused by serial MIDI data transmissions. We start with a set of onset times in seconds from note-on messages

$$M = \{x_1, x_2, x_i, ..., x_{|M|}\}. \tag{2.3}$$

Then we define a set of onsets N, such that

$$N = \{x_i | (x_i - x_{i-1}) > \delta\} \tag{2.4}$$

with $i = 1, 2, ..., |M|$ and $x_0 = 0$. $\delta = 0.07s$ is the threshold informed by the haptic timing limits of musicians. Musical events sounding below such a short period tend to fuse together into one musical entity. For the purpose of capturing expressive timing, rhythmic patterns, but also for the quantisation and transcription of rhythms, it is therefore useful to treat MIDI events falling below this threshold in such a way as if the would share exactly the same onset time. It follows from the above method that N will contain the unique onset times of single notes and polyphonic chord onsets. It is this set that we can analyse further using grouping, windowing and quantisation algorithms.

In order to extract note onset events from MIDI files we have used the BSD-licensed library *libSMF* by Eduard Tomasz Napierala[5].

MIDI in Real-Time

When receiving MIDI messages in real-time from a software layer, e.g., by using the cross-platform library PortMidi by Bencina and Burk (2001), all events are given a time-stamp in milliseconds. A millisecond timing resolution seems adequate because research pointed out that timing differences in the order of 1 - 1.5 ms are just audible, see Wright and Brandt (2001). In addition, PortMidi is a robust and reliable API used by many cross-platform open-source projects, such as Pure Data and Csound, as well as in one of the author's tools for algorithmic composition, see Boenn (2005).

2.3.3 Onset Data extracted from Audio Signals

Over recent years certain standards of audio feature extraction have emerged, like MPEG-7 for example, documented in Martínez (2004). An overview of signal processing methods for audio feature extraction is given in Peeters (2004) as part of the European CUIDADO project coordinated by IRCAM. Also, the *Mazurka Project* developed their own methods and tools, notably VAMP plug-ins written by Craig Stuart Sapp for the Sonic Visualiser. We also mention the Libextract library by Jamie Bullock that provides around 50 spectral and other features from audio data. The Aubio library by Brossier (2006) has mainly been used for our

[5] `http://sourceforge.net/projects/libsmf/files` [Accessed November 2010].

project in conjunction with the program Audacity for hand-editing. Onset detection still poses some problems notably when it comes to analysing vocal music and also string instruments. Sometimes small clusters of markers occur around a single audio onset or some of the audio onsets are not detected and markers are missing. Usually one adjusts the threshold parameter settings on the basis of test runs of the onset detection, but hand-editing of the resulting markers is still necessary if one needs to obtain a correct outcome of the tracking process. This is the case in our study with its main focus on quantisation and transcription of the collected onset markers. An in-depth survey of onset detection algorithms can be found in Collins (2006).

2.3.4 Is it sufficient just to know about the onset times?

One might ask why we can be so sure that onset data alone is sufficient for successful rhythm transcription. Is there enough information about the perceptual qualities of rhythm and metre encoded by a stream of inter-onset intervals? It might be useful to look at the psycho-acoustics of sound perception. The sound of musical instruments shows transients that are critical for the ability of humans to discern different musical instruments. Such transients are usually very short and happen during the attack time and so the note onset coincides with a very important physical and psychoacoustic feature of musical sound. It seems that the sound briefly occurring as the transient literally earmarks the moment in time when a new sound starts. One might think of a way how compound rhythms are perceived by a listener, namely through the connection of the individual attack points of different voices in the music.

Other researchers discovered a great sensibility of the nerve cells of the ear to detect onset and offset times of musical instruments playing a sequence of musical events. As pointed out in Brossier (2006):

> "Another type of temporal encoding operated by the human ear allows for the analysis of sonic events: some of the nerve cells are triggered at onset and offset times, where a sound starts and finishes [Whitfield, 1983]. The attack of the note, where the sound starts rising, triggers these nerve cells, while the sustained part of the sound, where the note is held, does not. A detailed analysis of this phenomenon was given in a study of the perceptual attack time [Gordon, 1984]. The study included perceptual tests in which listeners were asked, while listening to two different sounds separated by different time delays, to press a button if the attacks of both sounds were perceived as simultaneous. Gordon could measure accurately the perceptual attack times and found that these times were consistently different amongst different listeners. The tests showed that perceptual attack times of tones could be perceived significantly later than the physical onset of the sound in the music signal, up to a few tens of milliseconds, depending on the instrument played and the way it is played. Gordon [1984] observed that the perceptual attack time of several sounds was dependent on the timbre of the instrument, this quality of a sound which enables us to distinguish one instrument from another [Grey, 1975]."

Human perception of transients during attack time are crucial for the source identification of

musical sounds (Iverson and Krumhansl, 1993). Experiments where transients have been cut off have shown that instrumental timbre becomes hard to identify, this applies in particular to the family of wind instruments. This acute awareness of the onsets in perception also feeds back into instrumental practice. According to Kilian (2004): "Experiments showed that the onset time of notes usually are played more precise than their offset respectively duration."

Playing an instrument and listening to music depends heavily on certain temporal structures in our perception. There are musical phenomena of phrasing, expressive timing, a so-called dynamic function (Thakar, 1990), which describes the ebb and flow of tension in music as a whole, and there are phenomenal differences between musical performances of the same piece even when they happen under similar conditions. Central to all musical phenomena is the question of the temporal occurrence of sounds in relation to the sum of the effects of all musical events, which happened before, and in expectation of what will happen immediately afterwards as a direct consequence. A better understanding of the structure of human time perception is an important prerequisite in order to answer this question.

2.4 Temporal Perception

The following perceptual timing thresholds cited from London (2004) have important implications for the analysis of musical rhythm and metre. They will necessarily influence our computational model of musical rhythm in terms of the categorisation of rhythmic timing information, for example in algorithms that analyse sets of inter-onset times. However, one should treat the thresholds as guidelines not as absolute values. The physical timing values reported from various experiments have to be taken *cum grano salis*. There are inter-subjective differences, averaging procedures across subject groups and a mixture of experimental stimuli have been used, sometimes based on real musical performances using different styles, but often laboratory generated stimuli have been used. Nevertheless, these thresholds provide a highly valuable knowledge base for the analysis of real musical performances. For London (2004) the temporal limits provide a strong psychological foundation for the analysis of musical metres and their relationship with the performance and perception of musical rhythms.

Experiments have shown that human listeners spontaneously group identical and isochronous stimuli into groups of twos or threes. In other words, listeners are having a sense of accent on every second or third stimulus although such accents are not present in the stimulus. This effect of binary or ternary grouping is known as *subjective rhythmisation* (Bolton, 1894; Parncutt, 1994).

In the past, researchers have proposed certain ranges of perceptually relevant beat periods (London, 2004; Parncutt, 1994; Warren, 1993; Westergaard, 1975), which are remarkably similar. Following the thorough review of the research on tempo and beat perception by London (2004), musical beat periods are found in the range from 200 milliseconds (ms) up to 2 seconds, or 300 to 30 BPM. If one takes into account binary subdivisions at the fastest end, as well as ternary metres divided by the slowest beat, then one finds an outer timing envelope for metrical entrainment ranging from 100 ms up to 5 - 6 seconds. Within this range there is

another significant perceptual threshold relevant for the analysis and performance of musical rhythms. Parncutt (1994) had subjects tapping to a variety of rhythmic patterns. He found a peak of maximal pulse salience at 600 ms within a general range of pulse perception between 200 and 1800 ms. Durations between 600 - 700 ms are also known as *indifference intervals* (Fraisse, 1963; Wundt, 1911), which are neither overestimated (judged too long) nor underestimated (judged too short) by subjects in various studies. At 600 ms, on average, lies also the spontaneous tapping period (Fraisse, 1999) although this value seems to vary with the age of the subjects. London (2009) calls the period of 600 ms *tempo guisto*, by sheer coincidence it corresponds to 100 BPM. London (2009) also suggests that this period relates more strongly to the neurobiological nature of motor control and kinematic movement as in walking and running, clapping hands (Repp, 1987), and so forth, as opposed to metabolic periodicities, such as breathing and heart beat.

2.4.1 Shortest Timing Intervals

Hirsh (1959) found that a 2 ms separation is required in order to discern two tone onsets, otherwise they would be heard as a single one. But, if the listener is asked to report the order of the two onsets correctly, then they need to have at least a 20 ms interval between them.

It appears that fast played ornaments and gestures, such as trills, drum rolls, arpeggios or bow tremoli, fuse perceptually together into a single sound event. The time periods for this kind of musical gesture are approximately 70 to 90 milliseconds long. The upper limit of haptic frequencies is 11 Hz on average, which corresponds to a period of approximately 91 milliseconds (ms) (Verplank et al., 2000). Trills of professional piano players have been analysed by Brown and Smaragdis (2004). The highest frequencies can reach up to the range of 14 Hz (peak values could even reach 16 Hz), with a period of about 71 ms, i.e., the time elapsed between individual notes of the trill. It is interesting to note that such periods of oscillation stay just above the human limit of pitch perception: 16 Hz $\equiv$ 62.5 ms.

2.4.2 The 100 ms Threshold

London (2004) found that only above the 100 ms threshold elements of rhythm and metre can be recognised one at a time. "The lower limit for metre, that is, the shortest interval that we can hear or perform as an element of rhythmic figure, is about 100 milliseconds (ms)" (London, 2004, p.27). This threshold is consistent with music psychological research. For example, the lower limit for subjective rhythmisation is 115 ms (Bolton, 1894). The threshold for reliable durational discrimination lies also at 100 ms (Hirsh et al., 1990). It is the minimum period, which is necessary for the cortical processing of sound patterns (Roederer, 2008). It is also the fastest period of pulsation to which subjects could synchronise a four times slower tapping period (Repp et al., 2002). However, the average of his experiments peaks at 120 ms. 100 ms is the shortest duration in ride cymbal patterns of jazz drummers (Friberg and Sundström, 2002). The fact that 100 ms is the shortest possible metric subdivision (London, 2004) has consequences for the durations of performed metrical structures as we will see later.

2.4.3 Fastest Beats

Beats are the governing force within a metrical hierarchy. Given the minimum duration of 100 ms for any rhythmic element that has an identity of its own as part of a musical gesture and given the assumption that "hearing a beat requires at least the potential of hearing a subdivision", (London, 2004), the threshold for the fastest beat must be assumed to be in the range of 200 to 250 ms. This range correlates nicely with various results from music psychological experiments. It is the borderline period between holistic versus analytic processing of rhythmic events (Michon, 1964). It is supposed to be the limit for short-term memory (Crowder, 1993). It is the cutoff for backward masking (Massaro, 1970). In this range a shift happens in nature of the just noticeable difference (JND) for duration (Friberg and Sundberg, 1995). For Westergaard (1975), this is the range of the fastest beats: 240 to 300 BPM. Repp (2002b) found this rate as the limit of the subdivision benefit. There is a benefit for listeners to tap along at certain subharmonic periodicities, e.g., 1:2, 1:3 or 1:4, if the stimulus pulsation is sufficiently fast. Warren (1993) found here the upper limit for melody recognition. It is also the upper end of range of pulse salience (Parncutt, 1994).

2.4.4 Slowest Beats

On the other side of the beat spectrum we find the slowest tempo range of 40 to 30 BPM, or durations of 1500 to 2000 ms per beat (London, 2004; Westergaard, 1975; Warren, 1993). Here one finds a correlation with the upper limit of subjective rhythmisation (Bolton, 1894; Fraisse, 1999) and a shift from anticipatory attending to reaction time (Woodrow, 1932). The maximum time period for a ternary metre would then correspond to 4.5 to 6 seconds, which according to London (2004) is equal to the limit of psychological present (James, 1950; Pöppel, 1972; Michon, 1978; Fraisse, 1984).

2.4.5 The Perceptual Time Scale

Based on the perceptual timing thresholds presented by London (2004), we can now show them all together mapped to the basic musical time structures of beats, subdivisions and bars, see figure 2-9. Interestingly, the average indifference interval of 650 ms sits at the centre of the logarithmic scale of musical beat periods.

2.5 Agogics

The Oxford Companion to Music defines agogic accent as follows, after Efrati (1979, p.101):

> This term [...] is applied to that kind of accent which belongs to the nature of the phrase, as distinct from the regular pulsation of the so-many-beats in a measure, and which is produced rather by dwelling on a note than by giving it additional force. The first note of a phrase often suggests the desirability of a slight lingering, which constitutes an agogic accent. So does a note longer than, or higher or lower

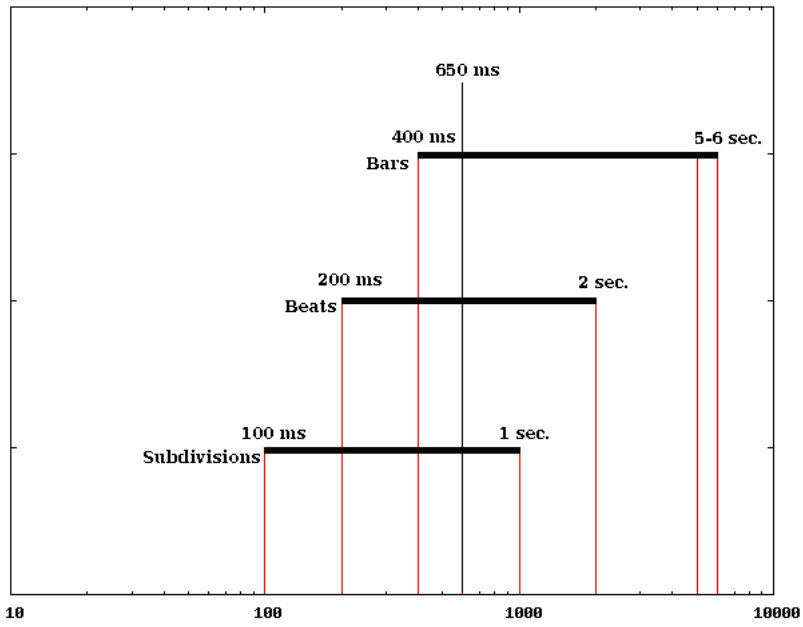


Figure 2-9: Perceptual timing thresholds and musical time structures.

> than, those that have preceded it. So does a pungent discord about to proceed to its resolution.

This notion of agogic accents proves again how Thakar's *dynamic function*[6], that of the gathering and dissipation of musical energy, permeates many different parameters. In our opinion, it is according to the *dynamic function* that musical agogics work although there hasn't been a systematic study yet to establish a link between those. Agogics are often related to rubato playing[7] for which there are two main historic traditions. The old tempo rubato meant to maintain the tempo of the musical accompaniment whereas the melodic voice, by contrast, was allowed to halt, to slow down or to speed up, thereby creating *notes inégales*. The resulting new relations of consonance and dissonance between the melody and the accompaniment were set intentionally and formed a part of an improvisational element where the musical texture was enriched by new appearances of dissonances and consonances. Of course those new relations needed to follow the common rules of counterpoint but they enriched the musical experience of a piece and added musical events not explicitly written down in the score, similar to the practice of adding ornamentation to a melody.

Theoretical works discussing this kind of rubato are Tosi, Quantz and Carl Philip Emanuel Bach.

> Slight tempo modification in the principle voice, for the sake of expression, was permitted. The so-called "tempo rubato" took its origin from seventeenth-century

[6] not to be confused with terminology of musical dynamics such as *forte* and *piano*.
[7] tempo rubato means literally 'stolen time'.

> Italian vocal music, and was first described by Tosi in his "Opinioni" (1723). His definition: "rubamento di Tempo, sul movimento uguale d'un Basso." Whilst the principal voice was allowed some tempo variation, the accompaniment has to carry on its regular pace.
>
> C.Ph.E.Bach (Chapter XXIX, §18): "When in a polyphonic composition the bass, as well as the ripieno parts, have to hold a long note, whilst the principal voice maintains its special movement, *even introduces some irregularities now and then;* the accompanist would do well, in order to maintain a strict time, to go on striking the beats with his right hand." (Efrati, 1979, pp.217)

The other form of tempo rubato refers to the notion of the complete set of voices changing their tempo according to the Gestalt rule of common fate (Deutsch, 1999a), thus not changing the synchronisation between different voices but changing the tempo as a means of expression to serve the *dynamic function* of the musical texture. Of course during the Romantic period where Chopin is often taken as an example, both forms of rubati could be mixed in that the melodic layer and the accompaniment both play rubato but at the same time they may also be slightly out of synchronisation and thereby creating additional harmonic and rhythmic events adding to the flow of musical tension and resolution.

There is an understanding of rubato that demands that the tempo changes only for a short period of time and then it returns back to a sort of average value applied to the entire piece. A rule attributed to Chopin[8] states that if within one bar one slows down for example, the next bar has to compensate by accelerating proportionally to the previous bar's rallentando. In this way performers try to level the amounts of rallentando and accelerando during a piece so that these tempo deviations always return like a flexible matter into its state of rest, i.e., to the overall average tempo of the performance. Rubato playing does not have total liberty but it is set into a context: "The detailed temporal properties of performed music [...are] often referred to as temporal micro-structure [...]". (Clarke, 1999) And it is a micro-structure set into relation to the structure of larger parts of the piece and finally into relation of the piece as an entity, therefore still aiming at a perception of unity.

The presence of rubato and expressive timing in musical performances means that automated tools for quantisation and transcription need to become aware of continuous tempo changes. Yet we have seen that tempo itself is a perceptual construct and not easy to extract from the actual onset data. Therefore we recognised that a Gestalt approach was needed in order to cope with rubato playing and expressive timing.

2.6 Musical Ornaments

This section is based on our published ICMC paper (Boenn, 2007a). We focus here on the problem of automatic quantisation and rhythmic transcription of syncopated rhythms and Baroque ornaments, e.g., appogiaturas, mordants and trills from time-tagged audio recordings

[8] Personal communication of the author with pianist and teacher Elisabeth Wangelin-Buschmann.

without knowing the score in advance. The technology described in our thesis has been tested with the transcription of the *Aria* of J. S. Bach's *Goldberg Variations*, BWV 988, recorded by Glenn Gould in 1955 and 1981[9].

We want to grasp the subtleties in rhythmic expression that cover the use of agogics (rubato, expressive timing), as well as the use of ornaments like the trill or mordant, which are to be played in relation to the musical context, neighbouring durations and the current tempo, but also offer room for tempo fluctuations and deviation from straightforward metrical playing. Certain manners of playing, which are not directly encoded in notation, can become characteristic of a musical style. The playing of over-dotted notes in Baroque music is a much debated example of such a practice (Efrati, 1979; Fabian, 2003). Finally, the fact that no musician plays metronomically exactly stems from the human need for musical expression, which can only materialise itself in time. As already mentioned, time-tagged data can be derived from audio or MIDI files or modelled by a performance editor[10]. Our rhythmic analysis of timing data takes into account that the tempo within a musical performance is in a state of constant flux. The two main problems are to detect at the same time tempo changes on different scales[11] and to identify the metrical relationships between all note-onsets (Raphael, 2001).

Musical ornaments are a particularly challenging problem. Trills are traditionally not written out note by note, their execution and timing leaving considerable room for interpretation and expression on the part of of the musician. The rapid interchange of two notes might also lead to the perception of a single sound event, where the individual notes fuse together creating an impression of unity, rather than a sequence of discrete note events. On the other hand, trills can also be played out slowly, especially within a calm and slow movement[12]. The challenge for any automated transcription system is that it should distinguish the ornament from other events and correctly identify and transcribe its onset locations. Trills appear in a great variety of stylistic flavours. In order to give some rules to the performers, many educational works of the late Baroque and early Classical eras indicate with metrical precision how a specific ornament should be executed (Efrati, 1979). However, scholarly works on ornamentations are not consistent (Fabian, 2003). More importantly, many players choose to execute ornaments freely in an improvised, spontaneous manner[13]. Therefore, a transcription algorithm cannot rely simply on pattern matching based upon an encoded metrical table of all 'historically approved' practices. The challenge is the rhythmic transcription of quantised ornaments and onset data collected from audio recordings. We do this first by investigating closely every single event in its rhythmic context and we will finally achieve the kind of notation that is added to critical editions explaining the execution of an ornament[14]. Pitch information and knowledge about

[9] discography at http://sonybmgmasterworks.com/arists/glenngould/

[10] e.g., Director musices by KTH

[11] Honing (2001) suggests that there are two main timing procedures, one related to global tempo changes like rubato playing and the other being independent of the global tempo but relating only to the onset location of the current beat, e.g., swing or grace notes.

[12] see Tureck, cited in Bazzana (1997, p.232)

[13] "[Hans] Klotz re-affirmed that the metre of ornaments was basically free, the notation only indicative (Klotz, 1984, p.37)."(Fabian, 2003)

[14] Gould studied the Kirkpatrick edition (1937)(Bazzana, 1997). Although this edition contains a transcription of the *Aria*'s ornamentations, Gould does not adhere to it by the letter.

the rules of counterpoint have to be added at a later stage, enabling the program to provide common practice shorthand notations. We are aware of previous research carried out in the areas of rhythm quantisation and tempo tracking (Cemgil and Kappen, 2003), beat tracking (Dixon, 2001; Hainsworth and Macloed, 2003) and automated rhythm transcription (Raphael, 2001). Although none of them addresses the problems arising from ornamentation practices, their general use for transcription of time-tagged audio data is acknowledged. Their models incorporate Bayesian statistics, hidden Markov models, Kalman filtering, neural networks and dynamic programming. The extraction of piano trills has been addressed in Brown and Smaragdis (2004). Their method uses a statistical analysis of spectral data from a database of trills only. There has been a recent approach to the detection of chord spreads (arpeggios) and trills within the MPEG-7 context (Casey and Crawford, 2004). In the latter it is assumed that ornamentations are always relatively fast sequences of notes, but this assumption is far too general to become a model for all trills and arpeggios no matter which musical context, let alone other kinds of musical ornament. They are not necessarily played 'as fast as possible'. On the contrary, if those more slowly improvised events can easily be confounded with other note durations written in the score, it is likely that a neural network trained with data from one pianist has to deal with unexpected events, when confronted with the ornamentation practice of another performer, which might lead to errors in the transcription.

We therefore propose a model that involves no machine learning or neural networking techniques. More importantly, our model does not only look at the immediate predecessor of an event as in the published methods using directed acyclic graphs (Cemgil and Kappen, 2003; Raphael, 2001), but it rather takes into account each and every connection, forwards and backwards in time, of a single event with all remaining note events within a given analysis window. In this respect our approach is closer to Dixon's clustering method used for beat tracking (Dixon, 2001).

The Bach *Aria* was chosen because it is a highly ornamented piece. There are two famous recordings by Glenn Gould that are completely different in style and tempo. The tempo of the latest recording of the *Aria* is about two times slower than in 1955. In addition, the *Aria* is repeated each time at the end of the cycle. Besides its rich ornamentations the *Aria* shows interesting balances between repetition and variation of rhythmic and melodic figures (Gestalten).

The time-tags were generated with the C program library Aubio[15] and have been edited manually with Audacity. A single list of onset data has been produced that represents the combined rhythm of both hands. Gould, like other Keyboard players, sometimes performs inequalities (Fabian, 2003) between both hands in order to accentuate a musical event. Thus we were able to explore the effect of independent part playing on our algorithm.

The kind of ornaments used by Bach are in order of their appearance: Mordant, Appogiatura, Double Cadence, Cadence, Trill, Arpeggio, Accent and Trill, and Slide. Gould adds in both recordings two inequalities in the alto voice (bars 16 and 24 on beat 2) to create additional 4-3 suspensions. In 1981 he adds an inequality in the bass line in bar 19, beat 3, in order

[15]see Brossier, P.M., http://aubio.piem.org

to accentuate the dissonance with the trill of the alto voice. He also resolves with inequality the appogiatura of the soprano in bar 26.

2.7 Gestalt Theory

An important part of our research relates to Gestalt-Psychology. Our clustering method of grouping IOIs into distinct duration classes, as described in chapter 5, depends on the notions of proximity, similarity, good continuation and on the principle of common fate. Durations that look identical in CPN are not exactly the same in a musical performance. In order to classify IOIs according to similarities in their temporal extension, we analyse the spatial information in graphical representations of time-dependent musical processes, see the use of the adjacent interval spectrum in section 2.2.4 and figure 2-8. In addition, the space-time correlations as expressed in language for example might point to the fact that time is often experienced in spatial terms as outlined by Köhler (1947). Proximity is a spatial concept after all, or to quote from Köhler's "Gestaltpsychology", pp. 88:

> "Experienced time has certain characteristics in common with experienced space, particularly with the spacial [sic] dimension which is indicated by the words "in front" and "behind". Words which refer to relations in this dimension are used as terms for temporal relations everywhere and in all languages. In English we may have something "before" or "behind" us both in the spacial [sic] and temporal meaning: we look "forward" in space as in time, and death approaches in time just as somebody approaches us in space. From the point of view of isomorphism, one would expect that there is a corresponding kinship between the physiological correlate of the temporal and that of this spatial dimension. At any rate, temporal "dots" form temporal groups just as simultaneously given dots tend to form groups in space. This holds for hearing and touch no less than it does for vision."

This passage can also be found in Efrati (1979, p.75), where he reports views on the musical line in the context of Bach's Solo Violin Sonatas and Partitas and the Suites for Solo Cello.

We will demonstrate that the "temporal dots" are equal to the given IOIs forming "temporal groups", which can be analysed by our clustering algorithm in chapter 5. It uses the concept of proximity, i.e., similar magnitudes of duration cause the formation of a temporal group. This is different from the traditional use of grouping that is applied to a structural sequence of musical events, for example in Lerdahl and Jackendoff (1996), where the aim is to find motive boundaries for example. We see a specific temporal group as a duration class and it is the ratios between different duration classes that form rhythmic motives, Gestalten, phrases, sections, and so on. The important point is that the duration classes are not fixed arrangements of magnitudes but rather that their mean values can vary significantly in the course of a performance. Although this happens during a musical performance the ratios between duration classes remain generally stable, a fact that is linked to the other Gestalt principles, such as good continuation and common fate. Here are those definitions by music psychologist Diana Deutsch (1999a, p.300):

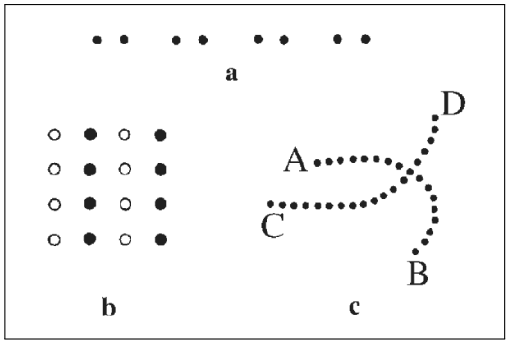


Figure 2-10: "Illustrations of the Gestalt principles of proximity, similarity, and good continuation." (Deutsch, 1999a, p.300, Figure 1). Reproduced with kind permission.

> The early Gestalt psychologists proposed that we group elements into configurations on the basis of various simple rules. (see, for example Wertheimer, 1923). One is proximity: closer elements are grouped together in preference to those that are spaced further apart. An example is shown in Figure [2-10a], where the closer dots are perceptually grouped together in pairs. Another is similarity: in viewing Figure [2-10b] we perceive one set of vertical rows formed by the filled circles and another formed by the unfilled circles. A third, good continuation, states that elements that follow each other in a given direction are perceptually linked together: we group the dots in Figure [2-10c] so as to form the two lines AB and CD. A fourth, common fate, states that elements that change in the same way are perceptually linked together. As a fifth principle, we tend to form groupings so as to perceive configurations that are familiar to us.

Our algorithm for rhythm analysis and transcription will exploit the above cognitive concepts. The fifth principle, to bias cognition of objects into ones that we already know, is related to learning. In fact the majority of the research in our field is concentrated on that particular principle of recognition of learned patterns. However, as we demonstrate in this thesis, the other four principles have been proven to be of even greater importance in guiding our development of an automated transcriber. We believe that the aspect of memory, a database of learned patterns, will come as a natural by-product of the cognition process. The successful matching of patterns can have an affirmative role, but we are not aware of a successful pattern extraction AND recognition program in the field of computerised rhythm recognition. In pattern-matching solutions such as Kilian (2004) or Bayesian techniques applied by Cemgil et al. (2000) and Raphael (2001), it is always the user who needs to provide the patterns in the first place, e.g., the machine needs to be trained in how a certain musical style or performance relates to a given score of that particular style. As far as we know there is no program yet that can automatically expand its knowledge base of patterns, e.g., learning a new style automatically. We think that our tool can be used in the context of pattern extraction, since our quantisation leads to results that are easy to analyse and to categorise. Categorisation of

rhythmic Gestalten is seen as an important feature of human rhythm perception, see Papadelis and Papanikolaou (2004), Desain and Honing (2003). There is also an interesting link with speech cognition:

> Among the difficulties faced by listeners are the speed of spoken language, the segmentation problem, co-articulation, and individual differences in speech patterns. Listeners cope by taking account of possible-word constraints, stress patterns within words, and the fact that co-articulation is generally greater within than between words. There is categorical perception of phonemes, but we can discriminate between sounds categorised as the same phoneme. Ambiguous phonemes are more likely to be assigned to a given phoneme category when that produces a word than when it does not, and this lexical identification shift seems to be perceptual. The phonemic restoration effect within words involves perceptually restored phonemes produced by top-down lexical activation rather than by pure guessing. (Eysenck and Keane, 2005, pp.357)

It would be interesting to see if there is a comparable "lexical shift" present in rhythm perception along the lines of the argument that ambiguous inter-onset ratios are more likely to be assigned to a given rhythm when that leads to a match with a previously learned rhythm category. Those rhythm categories would involve only simple integer inter-onset ratios. The works of Papadelis and Papanikolaou (2004) and Desain and Honing (2003) seem to support that idea. We also follow this notion so far as our process shifts complex inter-onset ratios towards simpler ones using a metric that we will describe in detail in chapter 6, pp.137. What we have not implemented yet is a comparison of quantisation results with learned patterns. This process could be used to reinforce the search for correct or preferred rendering solutions, which involve the transcription of the quantised stream of audio onsets. Yet, as we will demonstrate later, our system is also working without this extension. A database of rhythmic patterns would be useful in order to facilitate further analysis of the quantised data. A method could be devised that extends our method in order to automate building such a database on the basis of successful onset quantisation.

There is another aspect of Gestalt psychology that we took into account. The idea that there are no isolated musical events, that the term "participation in a line" expresses a deeply rooted musical and human concept, that of the individual tone taking part in an entity that means something more or rather different from all the meanings of the individual members summed together. We would like to give here two quotes from the same chapter in Efrati (1979):

> We feel a musical line to have significance when the tones group together to form a pattern, in the sense of Gestalt psychology. In this case every tone contributes to the expressivity of the whole, whilst each gets its own meaning from its position within the group, as well as from the form of the latter. (Efrati, 1979, pp.74)
>
> According to Christian von Ehrenfels a "Gestalt" is characterised by a) its supersummation (a melody signifies more than merely the sum of its tones, a face is

> more expressive than the sum of its constituent parts), and b) by the possibility of transposition. (Efrati, 1979, p.75)

We believe that in the musical domain the principle of supersummation has not been studied thoroughly enough and is not entirely understood. The important idea that we were able to successfully implement in our algorithm is the fact that every single event, inter-onset-intervals (IOI) in our case, is seen in relation to its many surrounding events thereby creating a dense network of inter-relations, which helps us to group clusters of similar IOIs together[16].

The Gestalt principle of proximity governs also the vertical axis of polyphonic ensemble practice and the perception of musical events that are supposed to start together at a certain onset time. But when scores indicate that events start together it does not necessarily mean that the onset times need to be exactly the same for these events. Physically, those onset times might be slightly different although players and listeners perceive them as being together.

> Rasch (1988) [...] made recordings of three different trio ensembles (string, reed, recorder) and calculated the onset relations between tones when they were nominally simultaneous. He found that asynchrony values ranged from 30 to 50 ms, with a mean asynchrony of 36 ms. Relating these findings to his earlier perceptual ones, Rasch concluded that such onset asynchronies enabled the listener to hear the simultaneous sounds as distinct from each other. According to this line of argument, such asynchronies should not be considered as performance failures, but rather as characteristics that are useful in enabling listeners to hear concurrent voices distinctly. Deutsch (1999a, p.306)

The physical imperfection of asynchrony between voices that should start together seems to facilitate polyphonic voice separation, which made possible the polyphonic complexity essential to Western music throughout large parts of its history. On the other hand, there is a tendency in recordings of popular music to eradicate this phenomenon by perfectly aligning beat onsets manually in the studio, but then there is not much complex polyphony happening in rock music. Especially dance music nowadays requires a high amount of 'tightness', i.e., events perfectly aligned to a beat and that beat kept in mechanical sample-accurate perfection. The notion of perfection in industrialised manufacturing processes seems to have inspired record label producers to 'tidy-up' any hint of so-called musical imperfection. This could also be an area where automated quantisation tools, such as ours, are useful: if they are adequately capable of maintaining graphical score-representations, they can be used as a clock-source to keep a sample-accurate timing of beats and can serve to align events automatically while maintaining a plausible musical structure. The method could be facilitated by using common time-stretching

[16] Influences of the principle of supersummation can be seen in various works of the 20th century avantgarde music. For example Pierre Boulez:

> If you look at a forest you can't see the individual leaves from the distance, but every leaf is there.

From an interview with Boulez with regard to the musical complexity and instrumental detail of his orchestral rewritings of his early piano work *notations* performed by the *Junge Deutsche Philharmonie* in 1991. Or see James Tenney (1992) for his account of 20th century music aesthetics, where Gestalt psychology plays a great part in an attempt to define a new language for music theory where the term *clang* plays a pivotal role

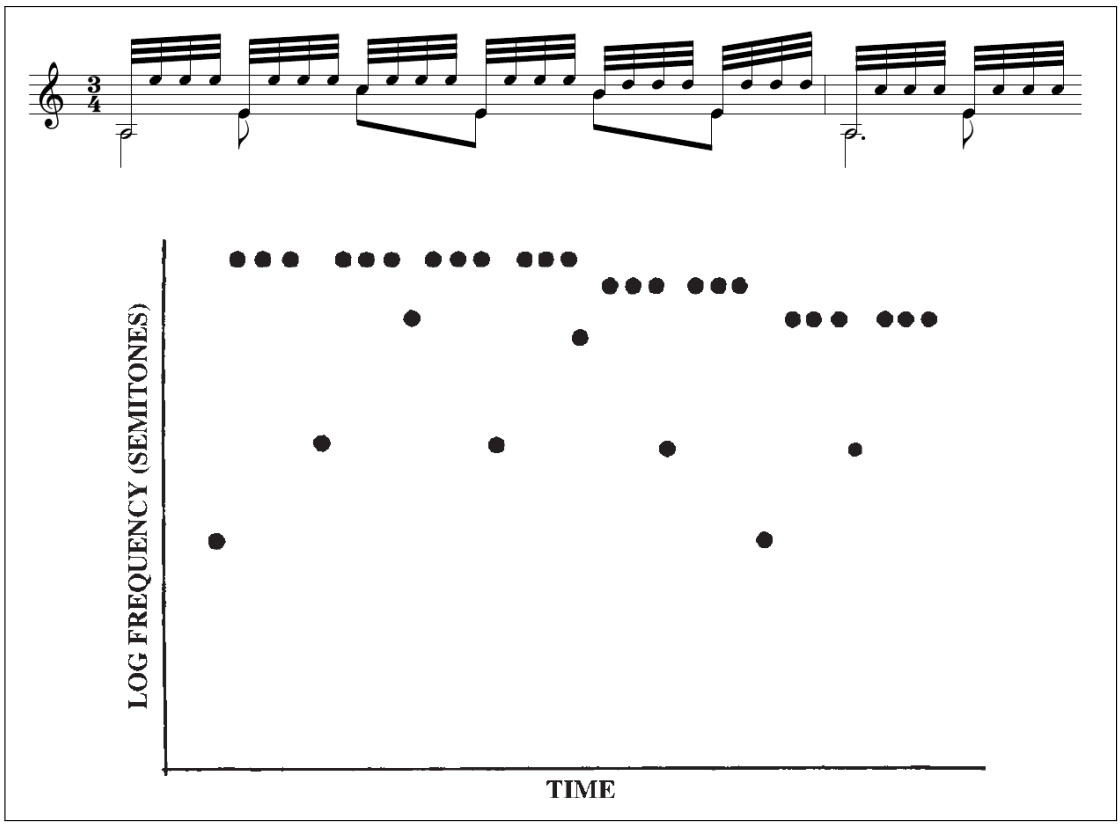


Figure 2-11: "The beginning of *Recuerdos de la Alhambra*, by Tarrega. Although the tones are presented one at a time, two parallel lines are perceived, organized in accordance with pitch proximity." (Deutsch, 1999a, p.309, Figure 5). Reproduced with kind permission.

tools in order to adjust the length of inter-onset audio fragments and make them fit into the quantised rhythmic structure.

Another example of the strong influence of grouping mechanisms is again given by Deutsch (1999a, p.309):

> In music played by [plucked instruments], when the same tone is rapidly repeated many times, and it is periodically omitted and replaced by a different tone, the listener may perceptually generate the omitted tone. Many examples of this phenomenon occur in 20th century guitar music, such as Tarrega's *Recuerdos de Alhambra*, shown in Figure [2-11], And Barrios' *Una Limosna por el Amor de Dios*. Here the strong expectations set up by the rapidly repeating notes cause the listener to "hear" these notes even when they are not being played.

The active (re-)construction of musical events that follow Gestalt principles also inspired our quantisation algorithm. Just like melodic movements being actively completed by the listener, we believe that beat induction and the recognition of rhythmic categories in spite of the underlying tempo being varied all the time, are all proactive achievements of the listener

who rather unconsciously simplifies the ratios between onsets in order to match certain rhythmic categories aligned to an underlying pulse structure. In a variation of Ockham's razor we suggest that if there are two or more rhythmic explanations of a series of inter-onset-intervals then the simpler explanation is the one that most likely represents the rhythmic sequence. Then there is the question of how we define simplicity. We will show how our program is capable of measuring various degrees of simplicity of a score representation on the basis of musical performances only and without knowing the score nor the style of music or performance in advance.

Finally, music psychology again draws attention to the fact that timing processes are having an effect on the perception of other musical parameters and *vice versa*. In relation to the idea of inter-connectivity between single note events, research has pointed to the fact of a necessary *temporal coherence* as a function of pitch proximity and tempo:

> The term temporal coherence is used to describe the perceptual impression of a connected series of tones. The conditions giving rise to temporal coherence were studied by Schouten (1962). He found that as the frequency separation between successive tones increased, it was necessary to reduce their presentation rate in order to maintain the impression of a connected series.
>
> Van Noorden (1975) examined this phenomenon in detail. Listeners were presented with sequences consisting of two tones in alteration, and they attempted either to hear temporal coherence or to hear *fission* (i.e., two streams of unrelated tones). Two boundaries were determined by these means. The first was defined as the threshold frequency separation as a function of tempo that was needed for the listener to hear the sequence as connected. The second established these values when the listener was attempting to hear fission. [...] When listeners were attempting to hear coherence, decreasing the tempo from 50 to 150 ms per tone increased the frequency separation within which coherence could be heard from 4 to 13 semitones. However, when the listeners were instead attempting to hear fission decreasing the tempo had little effect on performance. Between these two boundaries, there was a large region in which the listener could alter his listening strategy at will, and so hear either fission or coherence. So within this region, attentional set was important in determining how the sequence was perceived.
>
> Bregman and Bernstein (cited in Bregman, 1978) confirmed the interaction between frequency separation and tempo in judgements of temporal coherence. (Deutsch, 1999a, p.314)

This reminds us of our previous discussion on how harmonic complexity exerts a pressure within the vertical frequency-domain and on the horizontal time-domain, in so far as the perception of complex chromatic harmonies needs more time so that all aspects of voice-leading and all dynamic functions of tension and resolution are perceivable by the listener. If it goes too fast we lose out too many details and the evaluation of the dynamic functions of the sounds cannot take place. This is related also to the general fact that the frequency ratios between the sounds of a chord with increased harmonic tension are more complex than within a chord where

the individual sounds have simple frequency ratios, as for example in simple triads. A thorough discussion of the harmonic domain is however beyond the scope of this thesis. Readers might refer to Thakar (1990) for an introduction.

2.7.1 K-means Clustering

When being confronted with an ordered set of onsets from a musical performance, which we would like to transcribe automatically into CPN, the question is how we can order and classify the inter-onset intervals (IOIs) in such a way that they would represent specific duration symbols used in CPN. The solution we propose is to group onsets that are similar in length into a duration class. This is an application of the Gestalt principle of similarity. This grouping technique implemented in our quantisation algorithm is somewhat related to the well known K-means clustering method employed in unsupervised learning, see Jain and Dubes (1988) and MacKay (2003) for an overview. However, there are fundamental differences between the typical K-means algorithm and our approach. Our algorithm generates the clusters from the set of one-dimensional data alone and not by supplying an initial set of K randomly chosen points in order to start the calculation. One of the main questions with regard to traditional K-means clustering is: How do we determine the number K, i.e., how many clusters do we want or expect? Since the number of duration classes is unknown in advance but in the end has to be equivalent to the final number of clusters, we cannot determine the number K in advance, and an application of the K-means algorithm to our problem would have to guess K and perhaps do multiple runs, but then how do we choose the correct clustering from a series of calculations that gave us different results? See also the discussion in MacKay (2003, pp.287). These unknowns could lead to a potentially inefficient process, therefore we apply a deterministic approach using only the known IOIs to begin with. This method will be presented in chapter 5.

2.8 Metre

Metre can be described as an attentional dynamic process, whereby listeners actively anticipate future events using their individually entrained metrical framework, which is resonating to musical rhythms (Large and Jones, 1999; Large, 2001; London, 2004). Large (2001) describes metre as a self-organising dynamic structure based on a network of self-sustaining oscillations. Here, metre perception emerges from the formation of neural patterns stimulated by external musical rhythms. The oscillator-based entrainment model of Large and Palmer (2002) predicts the categorisation of temporally changing event intervals into discrete metrical categories. London (2004, p.161) regards metre as a form of entrainment, a "sympathetic resonance of our attention and motor behavior to temporal regularities in the environment." Entrainment enables listeners to extract temporal invariants from the music. The main structure of those invariants is encapsulated in the framework of metrical hierarchies and patterns of temporal expectancies that a human being learns when exposed to music over the course of his life.

2.8.1 Metre and the Dynamics of Attending

In many Western and non-Western musical practices, metre serves as the foundation for the performance and perception of rhythmic figures. But, metre is not just a neutral background structure, and its purpose is not only to enable the synchronisation of rhythmic gestures. Both metre and rhythm perception mutually influence each other. "Just as rhythmic patterns evoke metric responses to the listener, at the same time the listener's metric entrainment can give rise to a rhythmic figure" (London, 2004, p.58).

Metre is a "stable, recurring pattern of attentional energy". This pattern is represented as a cycle. Often metre is a metacycle, that is a set of hierarchically coordinated component cycles (London, 2004, p.66). Metre is a dynamic system which has various modes of excitation or resonance. The modes of resonances stem from different metrical levels and they form together a hierarchy of expectation (Large and Jones, 1999).

A musical metre very often contains more than one layer of pulsation (Barlow, 1991; London, 2004). Entrainment of metrical hierarchies can be modelled by Gaussians centred around metrical time points on different layers. The superposition of multiple layers will automatically create higher peaks for strong beats, the highest peak for the downbeat, and lower peaks will represent the weaker beats. The appearance of beats too early in a performance can then still be mapped to the expected beat position because of the continuous valued curve representing the level of expectancy that a listener is attuned to (Large and Palmer, 2002). While a weak beat has lost its influence on the current expectancy level the downbeat has already taken over and dominates the listener's expectancy within a certain window of time until a weak beat's expectancy rises again (see figure 3 in Large and Palmer, 2002, p.12).

In general, metre comprises a measure level starting with the downbeat, a tactus or beat level, which divides a bar into distinct timing regions, and often there are one or more levels of subdivisions per beat. There is no need for a distinction between metres and so-called hypermetres, which may span across many consecutive bars, because the number of metrical levels usually varies significantly during the course of a musical composition. The periodic change of attentional energy in metre perception leads to the assumption that metrical accents are generated by the listener. They are based on the strength and focus of entrained temporal expectancies (Barlow, 1991; Huron, 2006; Large and Jones, 1999; London, 2004). Music performers and composers can play with these expectancies, either to reinforce them, or to override them more or less gently.

The physical durations of metres, beats and subdivisions are not chosen by accident and in real performances they always arise out of the musical context. In addition, the perception of rhythm and metre, and thus their performance, is guided by certain periodicities and temporal limits. The temporal range for metric entrainment is between 100 ms and 5 or 6 seconds. A sense of beat is given in the range between 200-250 ms to 2 seconds. There is an attentional peak or preference for time spans around 600 - 700 ms, the indifference interval (see section 2.4). Those perceptual constraints lead to a specific grid of metrical and rhythmic possibilities. The following table 2.1 shows a tree structure that relates measures with different time signatures,

a beat level and its subdivisions to the psychological timing limits London (2004, p.44, Figure 2.6.).

duration or metre	value (float)	period [ms]	BPM	note name
beat period = 1200 ms, 50 BPM				
1/48	0.0208333	100	600	demisemiquaver triplet
1/36	0.0277778	133	450	
1/32	0.03125	150	400	demisemiquaver
1/24	0.0416667	200	300	semiquaver triplet
1/16	0.0625	300	200	semiquaver
1/12	0.0833333	400	150	quaver triplet
1/8	0.125	600	100	quaver
1/4 [beat]	**0.25**	**1200**	**50**	crotchet
2/4	0.5	2400	25	minim
3/4	0.75	3600	17	dotted minim
4/4	1	4800	13	semibreve
beat period = 650 ms, 92 BPM				
1/24	0.0416667	108	554	semiquaver triplet
1/16	0.0625	163	369	semiquaver
1/12	0.0833333	217	277	quaver triplet
1/8	0.125	325	185	quaver
1/4 [beat]	**0.25**	**650**	**92**	crotchet
2/4	0.5	1300	46	minim
3/4	0.75	1950	31	dotted minim
4/4	1	2600	23	semibreve
6/4	1.5	3900	15	dotted semibreve
8/4	2	5200	12	breve
9/4	2.25	5850	10	3 dotted minims
beat period = 430 ms, 140 BPM				
1/16	0.0625	108	558	semiquaver
1/12	0.0833333	143	419	quaver triplet
1/8	0.125	215	279	quaver
1/4 [beat]	**0.25**	**430**	**140**	crotchet
2/4	0.5	860	70	minim
3/4	0.75	1290	47	dotted minim
4/4	1	1720	35	semibreve
6/4	1.5	2580	23	dotted semibreve
8/4	2	3440	17	breve
9/4	2.25	3870	16	3 dotted minims
12/4	3	5160	12	dotted breve

Table 2.1: Three different metrical tempo grids after London (2004, p.44, Figure 2.6.). Tables show the beat periods for a crotchet (1/4 note) of 1.2 seconds, 650 and 430 milliseconds. Based upon psychological timing limits, three very specific sets of rhythmic subdivisions or metres emerge.

One can see that a chosen tempo for the central beat value automatically excludes certain note durations and lengths of metres because they exceed corresponding perceptual timing limits. Different tempi lead to specific sub-trees of metrical possibilities. Tempo changes on

the tactus level affect the perception of metre and these circumstances lead also to perception of groups and meta-groups in the performed rhythms.

There are certain rhythmic patterns that can theoretically fit in more than one metrical hierarchy. (London, 2004) calls them "metrically malleable patterns". But set into the context of a musical situation, there can only be one way of perceiving malleable patterns. Any pattern will be performed in a specific manner that unambiguously belongs to a specific metre due to the articulations employed by the performer. Musical articulation consists of phrasing, grouping, placing accents, and so forth, and it is metrically informed (Large and Palmer, 2002). It is a different question whether the composer might have used a different metre for the notation of the score, as compared to the musically performed and articulated metrical structure.

Polyrhythms, on the other hand, are created via specific integer ratios between note durations. Let us take for example the *hemiola*, two voices with durations in the ratio of 2:3, see figure 2-12. The resulting compound rhythmic structure will be played with a different

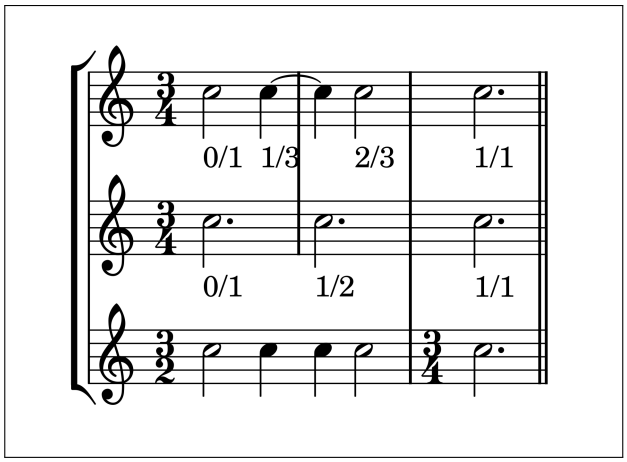


Figure 2-12: Hemiola with onsets marked by small integer ratios. The lowest staff shows its compound rhythmic structure.

metrical feel, because there can be now two simultaneous ways of articulation. The first one accentuates the onset at 1/2, the second will have the onsets 1/3 and 2/3 standing more in the foreground but still with the main accent on 0/1 and 1/1. One can produce polyrhythms easily within a metrical structure because of the basic isochronous pulsation within a metre. The system of CPN supports polyrhythms, because it is based on the concepts of metrical hierarchy. The interaction between metric patterns and rhythmic figures creates many possibilities for performers and composers to play with listener's expectations. Performers, for example, are able to phrase and place accents in contrast to the underlying metrical framework in order to communicate a specific musical meaning and to evoke specific listener responses.

London (2004) developed a circular representation of metre, similar to the necklace notation we have seen earlier, but extended through the use of subcycles, which reveal the different levels of a metrical hierarchy. This kind of representation underlines the cyclical nature of musical metre, which is grounded in the recurrence of patterns of attentional energy. At the lowest

level of a metre is the N-cycle, a set of N attentional pulses with the shortest period in a metre, but not shorter than 100 ms. N is always an integer so that there can be 8-cycles, 9-cycles and so forth. All higher levels of a metre, i.e., the subcycles, can be constructed by taking integer multiples of the period of the N-cycle, but as we will see, there are constraints for this procedure. The subcycle carrying the tactus or beat of the metre is called the beat-cycle. Listeners have a perceptual preference for beat periods between 500 to 700 ms. For N-cycles with N >= 4 there can be a special cycle marking approximately one half of a measure with an additional attentional period. The N-cycle and all subcycles viewed together form the metrical type, see figure 2-13. There are certain ranges for the tempo of a metrical type, where the periods of some individual subcycles align with the psychological timing limits mentioned before (see section 2.4). According to London (2004), these tempo ranges define so-called tempo-metrical types.

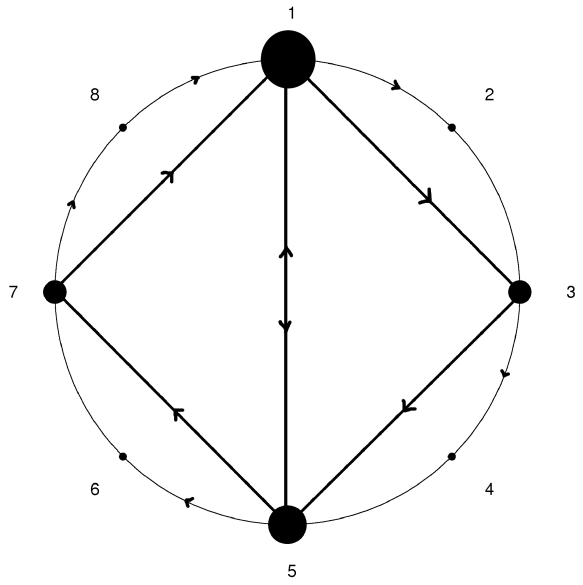


Figure 2-13: Diagram of a 4-beat 8-cycle isochronous metre including a half-measure level. The pattern starts at beat 1 with arrows indicating the direction of temporal flow. The size of the dots indicates different periods of pulsation. After Justin London, reproduced with kind permission.

This cyclical representation of metre is also useful for the visualisation of *well-formedness constraints* (WFC) (London, 2004). They are a set of general rules describing how a metre is constructed. as well as how the metre is heard, because they take into account the limits of temporal perception found in psychological research. The WFCs can be applied to metres in Western and non-Western music cultures. Their definitions are as follows:

> WFC 1: There are isochronous pulses of attentional peaks (time points) on the lowest level of the metre. This is the N-cycle. The period of pulsation must be >= 100 ms approximately.
>
> WFC 2: The N-cycle and all subcycles form a closed loop.

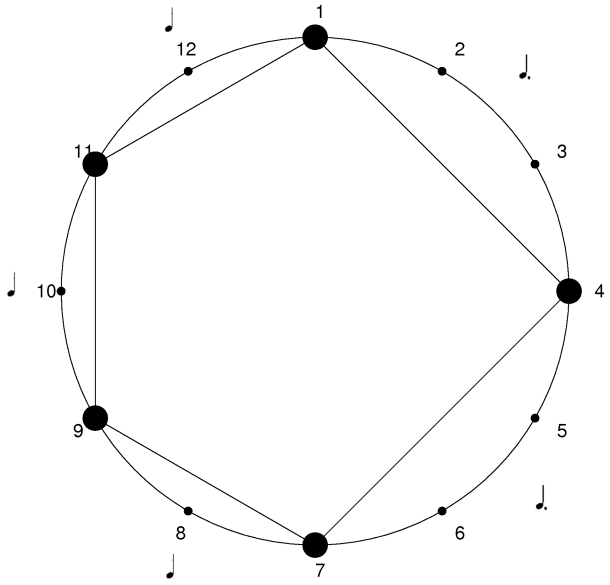


Figure 2-14: NI-meter structure 3-3-2-2-2 with 5 beats modelling Leonard Bernstein's "America" rhythm.

> WFC 3: Each cycle starts at the same time, at the same temporal location, and remains always in phase.
>
> WFC 4: Each cycle has the same cumulative period with a maximum of approx. 5 seconds.
>
> WFC 5: Subcycles connect time points of attentional peaks on the next lowest level, but not adjacent time points, i.e., they must skip over at least one time point.

The above WFCs have been developed on the basis of metres that exhibit isochronous beat cycles (I metres). But there exist also metres with non-isochronous beat cycles (NI metres), for example figure 2-14. London (2004) adds a sixth WFC of maximal evenness, so that well-formedness is also achieved for NI metres:

> WFC 6: The metre as a whole must be as maximally even as possible. Individual subcycles need not be perfectly maximally even; they may deviate form maximal evenness provided this deviation does not produce ambiguities or contradictions on the non-maximally even subcycle.

For London (2004), metrical hierarchies and patterns of expectations are not solidified but temporally elastic structures, which exhibit variance at different time scales due to the use of agogics in musical performances. The *Many Meters Hypothesis* states that "a listener's metric competence resides in her or his knowledge of a very large number of context-specific metrical timing patterns. The number and degree of individuation among these patterns increases with age, training, and degree of musical enculturation" (London, 2004, p.153).

2.8.2 Modelling of Neural Oscillations

Studies of weakly connected neural networks (WCNN) are interested in resonance phenomena between firing neurons (Hoppensteadt and Izhikevich, 1997). WCNNS and are reported to be useful especially for modelling pattern memorisation and recognition tasks of the human brain. They employ a network of periodically spiking neurons and connection functions. In such a network, pairs of neurons are modelled as pairs of coupled oscillators. "Interaction between each pair of oscillators averages to zero unless their frequencies are commensurable" (Hoppensteadt and Izhikevich, 1997, p.IX). Commensurabilty is established by a pair of frequencies whose ratio a/b uses relatively small and coprime integers a, b.

Large (2008) proposes a nonlinear system as a model for the neural oscillation of excitor-inhibitor cells, which allows him to develop a resonance theory of musical rhythm. Its goal is to predict the main psychological attributes of pulse and metre, which are not recognised as stimulus properties but rather as "endogenous dynamic temporal referents that shape experiences of musical rhythms".

Despite expressive tempo changes during a musical performance, metric pulsation is a musically stable phenomenon "in the sense that it can continue in the absence of a stimulus, and it possesses a form of temporal, or phase stability such that it normally synchronizes with events, but can persist in the face of rhythmic conflict, or syncopation" (Large, 2008, p.9).

Pulses can exhibit harmonic (2:1, 3:1, etc.) or subharmonic structures (1:2, 3:1), which align with the rhythmic and metric structures of a musical performance (Lerdahl and Jackendoff, 1996; London, 2004).

"The neural resonance theory of pulse and meter holds that listeners experience dynamic temporal patterns, and hear musical events in relation to these patterns, because they are intrinsic to the physics of the neural systems involved in perceiving, attending and responding to auditory stimuli" (Large, 2008, p.29)

Many experiments with artificial pulse sequences have shown that human motor behaviour can synchronise with periodic stimuli and adapt to phase and tempo deviations (Large and Palmer, 2002; Repp, 2002a, 2003). Moreover, stimulus events in periodic sequences can be anticipated (Repp, 2005). Pulse synchronisation is also maintained in the presence of syncopation (anti-synchrony), and against more complex rhythmic information within the stimulus. Large (2008) proposes a canonical model of the actual physics of neural oscillation. The following differential equation 2.5 of the complex variable z with respect to time is derived from the Wilson-Cowan model (see Large, 2008, p.19).

$$\frac{dz}{dt} = z(\alpha + i\omega + (\beta + i\delta)|z|^2) + cs(t) + h.o.t.) \tag{2.5}$$

with α as the bifurcation parameter, β is the nonlinear saturation parameter, ω is the natural frequency of the oscillator, $\omega = 2\pi f$ and f in Hz, and with δ as the frequency detuning parameter. The connection strength c controls the influence of the time-varying stimulus $s(t)$ on the oscillator. Higher order terms, $h.o.t.$, are ignored here for simplicity of the representation. This model can also be written in polar coordinates using $z = re^{i\phi}$ and Euler's formula $e^{i\phi} =$

$\cos\phi + i\sin\phi$:

$$\frac{dr}{dt} = r(\alpha + \beta r^2) + cs(t)\cos\phi + h.o.t.$$
$$\frac{d\phi}{dt} = \omega + \delta r^2 - c\frac{s(t)}{r}\sin\phi + h.o.t. \tag{2.6}$$

It is reported that the above canonical model captures certain universal properties of many known models of neural oscillations.

The mathematical analysis of the physics of neural oscillation makes several predictions in line with the resonance theory of pulse and metre. Endogenous periodicity is predicted by the spontaneous oscillations in the neural system. The generalised synchrony of pulse and metre with musical rhythms is predicted by the process of entrainment of nonlinear oscillations to an external stimulus. Finally, the perception of metrical accent is predicted by higher order resonances in nonlinear oscillators (Large, 2008).

When applying the above model of neural oscillation to the analysis of performed musical rhythms, the following questions arise: When does a given pattern of note onsets fall under the categories of pulse or metre, and when does the system recognise that a pattern falls outside of those categories? There is a theoretical possibility of using chaotic oscillations and/or burst oscillations in order to understand the perception and generation of complex (non-periodic) rhythmic patterns in music (Large, 2008).

However, in order to match the time scales of musical pulse and metre, it is not sufficient to use models based on single-neuron activity or alternating excitory and inhibitory neural activities. There is evidence that burst oscillation fits indeed well with musical time scales and this line of research is currently receiving much attention (Large, 2008; Large et al., 2010).

As an important expansion of the neural resonance theory of musical pulse and metre, Large et al. (2010) derived a canonical model for gradient frequency neural networks (GFNN) responding to time-varying external stimuli. The new experimental outcomes give supporting evidence for the resonance theory proposed by Large (2008).

2.8.3 Bayesian Techniques for Metre Detection

Temperley (2007) specifically addresses the problem of automated metre detection. He criticises Cemgil's and Raphael's works on quantisation and tempo tracking, as described in Cemgil et al. (2000), and Raphael (2001). Temperley states that both models implicitly find the correct metre by quantisation, they do not model metrical grids explicitly. Cemgil's model is reported to be only capable of dealing with simple rhythmic structures from performances. Cemgil's and Raphael's models involve a large number of parameters. But in all fairness, as one can see from Temperley's own description later, his method also works on a large number of parameters modelled by probabilistic tables that need to be learned from performances and score-representations of the Essen corpus of folksongs (Schaffrath, 1995). Also, the folksong melodies in that collection clearly show only simple rhythmic structures. In the end, Temperley admits that "the best way of modelling rhythm perception probabilistically remains to be

determined, and may well combine ideas from all of these models. (p.31)." His method tries to infer the metrical grid from the performed pattern of onsets according to Bayes' theorem:

$$P(grid|onsetpattern) = P(onsetpattern|grid)P(grid) \tag{2.7}$$

He uses a generative model of metre perception to calculate the most probable metrical grid to match the observed rhythm. At the lowest level his model has a fixed grid resolution of 50 ms called "pips". The system tries to map a three-level grid based on units of pips to the performance data: A tactus level #2 is tracking the beats, a sub-tactus level #1 tracks the downbeats, and the highest level #3 models first subdivisions of the beat. The random variables needed for the modelling are derived from the Essen corpus of folksongs. A tactus interval distribution is based on experiments in music psychology (Parncutt, 1994) that a stable tactus interval is approx. centred around 700 ms[17]. Temperley is aware that calculating the most probable grid on the basis on all possible grids would take a huge amount of processing time because of the number of possible grids and parameter combinations:

$$\begin{aligned} P(onsetpattern|grid)P(grid) = P(UT) \times P(LT) \times P(UPh) \times P(T_1) \\ \times \prod_{2}^{t} P(A_n) \times \prod_{2}^{t-1} P(T_n|T_{n-1}) \times \prod_{1}^{t-1} P(DB_n) \times \prod_{1}^{q} P(N_p) \end{aligned} \tag{2.8}$$

where $P(UT)$ is the probability of either a duple or triple upper level (subdivisions of the tactus), $P(LT)$is the probability of either a duple or triple lower level (downbeats), $P(UPh)$ is the probability of the phase of the upper level in relation to the tactus level (beats). $P(T_1)$ is the probability of the initial tactus interval indicating the initial tempo. This distribution prefers initial values of around 550 ms (109 BPM). Values above 700 ms, which is everything slower than approx. 86 BPM, have a combined probability of 0.01. The fastest possible initial tempo is modelled at approx. 133 BPM with a probability of 0.1. These tempo distributions are obviously style-dependent. $P(A_n)$ is the probability of generating another tactus beat, $P(T_n|T_{n-1})$ is the probability distribution of non-initial tactus intervals that is Gaussian-centred around 700 ms (86 BPM). Tempi faster than 133 BPM and tempi slower than 55 BPM are just not possible. $P(DB_n)$ is the probability of the position of each lower-level beat (downbeat) in relation to the previous tactus beat. If the tactus beat is triple, then there are two probabilistic variables for each tactus beat instead. $P(N_p)$ represents the probabilities of note-onsets at pips in relation to the pip's beat level.

In light of intractable solutions to equation 2.8, a dynamic programming strategy is applied that concentrates on finding the different tactus levels in a stream of onset data mapped to the underlying grid of 50 ms pips. Unfortunately, Temperley's explanation of his dynamic programming model is not always clear. For example he does not demonstrate the various initial states his system has to assume, for example is the first onset on the beat and at which metrical position? Temperley also does not mention that there is an implicit quantisation to a

[17]The 700 ms tactus interval corresponds to approx. 86 BPM

50 ms grid taking place, and we are given no explanation for the reason why he chooses 50 ms in the first place. A 50 ms difference between two onsets is audible, so we assume that rounding-errors from quantisation can possibly influence his results. Nevertheless, he shows an example of a successful 4/4 metre detection of one of the Essen folksongs that begins with an upbeat. Temperley tested his model trained with the Essen corpus with a subset of the corpus that was not involved in the training stage. The results show that his model is good at the beat-tracking level, the lowest metrical level, but less successful (at about 70 %) with detecting the higher metrical levels, e.g., downbeats. A comparison with a previously developed 'Melisma' model shows that the older system is more successful than the Bayesian approach, with an overall score of 0.793 for the new system compared to 0.865 for the 'Melisma' model. The reason for its better performance is likely to be the consideration of actual note lengths for beat detection. The fact that the system has been trained on a folksong database leads to question its ability to detect metre accurately on the basis of more complex rhythmic structure, e.g., syncopation, higher prime numbers in the number of beats per bar, for example in Eastern European folk music (see Bartók's essay *The So-called Bulgarian Rhythm* edited by Suchoff (1993, pp.40)), and higher prime numbers in the subdivisions of beats. Temperley's argument is of course the same for all machine-learning approaches in that the system just needs to be trained for those different circumstances. But when looking at the system's description, the probability tables for different complexities of musical situations are assumed to be fixed for an entire piece of music in a particular style. The system would rather need to process the information in parallel threads while analysing a piece of music in order to switch its probability tables to a different style when needed. We imagine that there is a lot of research effort involved to train the algorithm in a sufficient number and variety of styles, then also to devise a parallel computing listening algorithm that would be able to obtain enough data fast enough to let the system decide which probability tables are most likely to fit the actual musical information presented to the system and to provide sufficiently deep backtracking to recover from processing "in the wrong style". This probably could involve similar Bayesian techniques and dynamic programming strategies as presented by Temperley (2007). Although we assume that such parallel algorithms could be developed with modern cpu architectures, we think that the fine-tuning of this complex probabilistic system demands high amounts of time in research and testing its various models and outcomes.

Other problems reported in Temperley (2007) related to assuming a wrong metre, or detecting the right metre but finding the beats in the wrong places (phase alignment). It is a common argument that incorporating other musical rhythms, e.g., on harmonic or dynamic levels, might improve a system for rhythm detection that relies solely on note onset data. Yet, taking the note duration into account as for example in Raphael (2001), would require "a fundamental reworking of the system", admits Temperley (2007, p.47).

2.9 Quantisation

Quantisation is the mapping of performed note onset times and their durations to the onsets and durations defined by the score in CPN, or defined by learned metrical patterns.

2.9.1 Grid Quantisation

The task of quantisation, as understood traditionally in the context of Sequencers and Notation Software, used to be quite simple. An arbitrary stream of performed note onsets and end-of-note onsets is placed over a predefined equidistant grid of points on a timeline. This grid is defined, for example, by the length of a 16th note, given a certain tempo in BPM. Other grid resolutions can be set by the user, for example 16th triplets or 32nd note lengths. Onsets are then placed on the arithmetically nearest point on that grid. There are various grave problems that arise from this strategy. First of all there is the fact that the grid has no knowledge incorporated about the compound rhythms arising from using multiple subdivisions at the same time (polyrhythm). Even in its simplest form, i.e., 3 over 2, this knowledge is not represented appropriately by an equidistant grid. As a consequence, the proper differentiation between the beat division of 3 against 4 is often not possible. With simple grid quantisation, tempo changes during a performance are not tracked, therefore the quantised and transcribed result of an expressive performance soon gets out of synchronisation with the beat structure of the piece, nor are durations transcribed correctly, which leads to an unnecessary complexity of the transcribed results - the transcribed score as a consequence being cluttered with slurs between single notes and smallest note durations connected to long ones. In order to gain some precision with grid quantisation the user of a sequencer can record the performance together with a metronome click. But since real musical performances are not played against a metronome this method is of very limited use and expressive timings are ruled out from such a performance. It also happens that users might fall in front or behind the metronome click which again deteriorates the transcribed result. Because compound IOIs are not analysed, the grid quantisation method can lead to false note overlaps, i.e., note events are placed together as a chord in spite of having been played one after the other during the performance. The screenshots of figures B-1 and B-2 in Appendix B, pp.184, show our tests to play Bach's *Aria* of the *Goldberg Variations* (just the melody line with its ornamentations) using two commercial software programs, by the time of writing in their latest versions. Performances were recorded using the built-in metronome clicks.

2.9.2 Context-free Grammar

Longuet-Higgins and Lee (1982) proposed a method that infers a context-free grammar from a performed series of IOIs by first building IOI ratios, initialising beat duration based upon the first IOI ratios and by applying a small set of rules to keep track of changing tempi in comparison with the current tempo estimate. The rules are Initialise, Stretch, Update, Conflate and Confirm. A review of this method in comparison with other rule-based approaches can be

found in Desain and Honing (1999).

Context-free grammar and their application to music analysis was highly popular during the eighties, due to Noam Chomsky's work on the structure of the grammar of language. His main influence is evident in the GTTM by Lerdahl and Jackendoff (1996) and has been also popularised through Leonard Bernstein's TV production of his Norton Lectures (Bernstein, 1976).

2.9.3 Pattern-Based Quantisation

Kilian (2004) uses a pattern-matching approach for mapping IOIs from performances to score positions in CPN. In addition, a quantisation process takes into account the local context of IOIs that works on single voice streams including chords but without overlapping notes, i.e., compound rhythms that include chord onsets. He proposes a model that determines different levels of attraction between an unquantised onset and surrounding grid time positions. Kilian states that it is not always the closest grid position that forms the correct quantised time position within the score. Therefore he applies another weighting strategy that is built from the number of common time positions between different levels of subdivision of the metrical grid, for example $\frac{1}{2}$ coincides with $\frac{2}{4}$, $\frac{4}{8}$, etcetera. To account for groups of onsets for which no pattern can be found in the database, those gaps are recognised automatically and a history-based multi-grid approach is applied on their content. The histogram informs a weighting of simple IOIs based on previous IOI mappings (see also the tempo detection method described in 2.10.3 on page 71). The examples of a pattern database provided in Kilian's Appendix contain exact score solutions for entire bars and a few more different patterns. This obviously requires a preselection from the manifold of compound rhythms that can occur within one bar. Unfortunately, Kilian gives no indications on how to build the pattern databases in advance other than that the user is enabled to change the database content at any time. It suggests that the database needs to be adapted for different pieces of music and for different styles of music as well. There are no measurements provided as to what amount of information is needed by the program in advance in relation to the original score in order to perform well. Kilian also states that his model is an implicit Hidden Markov Model when he refers to research carried out by Takeda et al. (2003) who worked with learned patterns to infer score time positions and durations from performances. The pattern database for the quantisation task needs to include more complex patterns as compared to the database used for tempo detection.

Kilian presents successful quantisations of a Bach Minuet, first 16 bars, including a problem with transcribing an ornament[18]. He also shows his system's output of the first 16 bars from Beethoven's simple Piano Sonata Nr.20, the first movement of op. 42,2. The metre is $\frac{4}{4}$ and contains binary and ternary pulsations (quavers and quaver triplets). In addition, he provides a transcription of the melody part of Dave Brubeck's *Take Five* with correct swing rhythm transcription.

In general, Kilian states that the complex rhythmic patterns of the score as presented in the

[18] in the transcription Kilian uses the wrong sign (trill) for the performed mordant

examples above cannot be transcribed from performance data by using non-pattern techniques. That is precisely that what we are able to show with our approach in chapter 6.

2.9.4 Models using Bayesian Statistics

Raphael (2001) and Cemgil et al. (2000) report a system for note onset quantisation based on Bayesian inference. A performance rendering of a musical score delivers note onsets y whose positions on the time-line show a characteristic deviation from the score onsets x. Those deviations are due to a set of hidden variables θ, which can be trained with machine-learning techniques and therefore y can be traced back to the most likely score representation:

$$p(x, \theta|y) = \frac{1}{p(y)} p(y|\theta, x)p(x, \theta) \tag{2.9}$$

Both systems need to be trained with measured and analysed data from previous performances of the score that render a probability distribution of hidden variables θ. In addition, it needs to be trained with a probability distribution of onset occurrences within the score itself, i.e., when certain onsets are more likely to match with specific locations on a pre-defined metrical grid. This is dependent on style and genre and can vary even between pieces by the same composer or performer. The training stage needs to learn hidden parameters θ from previous performances that together with the score onsets prior and the observed onsets from the actual performance lead to an optimum score representation, i.e., as close as possible to the original score. It is from this learning process of the hidden variables where the difficulties of the system arise. There are integrations of the hidden variables required that are intractable in most cases. Therefore it is required to learn the best parameter settings from given observations, which is formulated as a maximum a-posteriori estimate (Cemgil, 2004):

$$\theta^* = \underset{x}{\operatorname{argmax}} \int dx p(x, \theta|y) \tag{2.10}$$

$$p(x|y) \approx p(x, \theta^*|y) \tag{2.11}$$

There are other areas in Cemgil's framework where he also reports intractable situations, generally solved by approximation strategies. It is also not clear how the choice of the underlying metre of a performance is automated.

2.9.5 IRCAM's KANT

The ICMC paper by Agon et al. (1994) describes KANT, an offline system for quantification of onsets and a library for for the Common-Lisp based software PatchWork. This program, which nowadays comes under the name OpenMusic, is a member of the family of Computer-Assisted Composition programs (CAC), where users can generated CPN scores via application of various algorithmic strategies, as for example proposed by Barlow (1984), Essl (1996) or Xenakis (1992). KANT is primarily aimed to facilitate the transcription of algorithmically generated rhythms and metrical structures. KANT is based on segmenting onset streams into measures,

followed by an approximate calculation of the underlying beat-length and finally quantisation of the onsets according to a user-supplied subdivision-grid mapped to the calculated beats. Segmentation is a semi-automated process via low-pass filter smoothing of the input function of discrete event durations and subsequent placement of "archi-measure" barlines according to the second derivative of the smoothed function of event durations. Further segmentations of the archi-measures are based upon an event-density calculation where maxima are being used to trigger a further subdivision into bars. The authors say that these segmentation processes still need a certain amount of hand-editing of bar-line positions in order to achieve good results. The quantisation stage works per beat of every measure and uses a least distance error strategy in order to move the original onset(s) to a position on a subdivision grid. Possible subdivisions s of the beat are in [1...32], with $s \in \mathbb{N}$. The set of quantised onsets is pre-ordered according to the least number of onsets eliminated. One of the problems with grid quantisation is that separately played onsets in the original stream can be forced to share the same position on the grid, also known as false-overlaps. Through the ordering of the results KANT favours quantisations with the least number of onset eliminations. False overlaps are generally avoided by resolving the issue of eliminated onsets through the use of grace notes. Then, after the first ordering, two different cost functions are being applied in order to create two versions of the list of quantised solutions within the frame of one beat. The first version is ordered according to the smallest distance given by the sum of all cubic ratios of the original durations over the quantised durations. The second list is ordered by an unspecified complexity measure for the chosen beat subdivision. The authors only state the order of complexity given by the first eight possible subdivisions:

$$1, 2, 4, 3, 6, 8, 5, 7, etc.$$

That particular order of integers came from discussions with one of the authors, Joshua Fineberg. According to personal communication with Gérard Assayag, the idea is that rhythmic complexity does not follow the natural order of integers. The subdivision of a beat by relatively high prime numbers produces a more complicated score. Other contributing factors to this kind of complexity are the use of nested subdivisions and the subsequent possibility of compound ratios by using slurs between two different note durations. A beat subdivision by 1, 2 or 4 is simpler than a division by 3, because in an otherwise binary rhythm context, a triplet will be used to give a particular effect, for example a ritardando, or the triplets employed in Jazz swing. Beat subdivisions 5 and 7 are even more uncommon. A different ordering of this list might be used if the music is structured by ternary metres and rhythms. The binary subdivision by 8 has a high complexity rank because the authors do not want to get a transcription with many 32nd notes unless necessary. Because of a small onset deviation, KANT could propose a complicated transcription with many linked 32nd notes, whereas KANT would produce a very neat 16th note transcription by ignoring this small deviation[19], hence the higher rank of the subdivision by 8.

The first solutions from both lists are then weighted by a user-supplied "precision" factor, which multiplies a Euclidean distance measure that compares both solutions with the original

[19] According to Assayag via personal communication with the author in February 2010.

onset stream. After the weighting, the solution with the least Euclidean distance is chosen for rendering. KANT treats all onsets falling into inter-beat intervals in this way.

The authors do not employ any beat or downbeat detection algorithms but rather rely on onset density distributions in order to group onsets into bars of unequal lengths - a strategy that serves certain aesthetics of composition and notation in 20th century avantgarde style. The approximative gcd algorithm is borrowed from solving the problem of detecting a fundamental frequency in audio signal analysis. The lengths of the measures are the input frequencies with the calculated beat-length as their fundamental. It follows that although bar-lengths are allowed to change, the beat-length is still considered to be constant and thus the underlying tempo is not allowed to change significantly. Expressive timing is only taken into account during the quantisation at beat-level. KANT requires a significant amount of user intervention at every stage of its processing, including editing archi-measures and bar-lines and decisions about the weights of cost-functions employed at the beat-level quantisation.

2.10 Tempo Tracking

Beat and tempo tracking are methods to measure instantaneous tempo in a musical performance on the condition that the music has an underlying pulse that is varied through expressive timing, stylistic conventions and musical necessity. Beat tracking is a special case of tempo detection combined with score following insofar as its output is a stream of pulses coinciding with the inferred metrical structure on the highest level, whereas tempo tracking does not return the beat structure directly but through the process of quantisation. Any successful quantisation of onset data has to track the tempo changes involved in an expressive performance. Brute-force quantisation algorithms on the other hand, like for example simple grid-quantisation to the nearest 16th note, are not able to take those tempo deviations into account and are therefore generally unsuccessful when it comes to human performances.

2.10.1 Multi-Agent Systems

The system presented by Dixon (2001) is capable of estimating the tempo and position of beats by processing MIDI and audio files. The input can follow an arbitrary musical style but is required to have a continuous underlying beat. Dixon's method takes the duration as well as the intensity of IOIs into account. Events falling under a 70 ms threshold are merged into a single event. The number of merged onsets adds then an additional weight to the event. From this 'cleared' list of events Dixon clusters IOIs together using a 25 ms tolerance for an IOI's distance from the mean of the cluster. In addition, the method also clusters together compound IOIs. The clusters can merge if their mean values become close enough during tracking. The means are allowed to shift during the performance. If a new event lies within 25 ms distance from the mean, the respective cluster accepts this event, which in turn increases its score. Clusters whose means are integer multiples of each other receive an additional weight $f(d)$

according to the integer factor d:

$$f(d) = \begin{cases} 6-d, & 1 \leq d \leq 4 \\ 1, & 5 \leq d \leq 8 \\ 0, & \text{otherwise} \end{cases}$$

The clusters with the highest overall score are used for establishing a current tempo hypothesis. Each hypothesis is then given to an agent. The states of the agents are then continuously checked to see if they would describe a viable tempo prediction. An event may fall outside the inner window of an agent into an outer window, in which case the current tempo hypothesis is updated according to the distance of the new event from the currently estimated beat position.

2.10.2 Probabilistic Methods

Documentation of the research in this direction can be found in Cemgil et al. (2000), Cemgil (2004), Cemgil and Kappen (2003). There it is reported that both, tempo detection and quantisation can be formulated within one and the same model described as a maximum a posteriori (MAP) state estimation task. Sequential Monte-Carlo integration methods are used for finding solutions.

Bayesian statistical methods seek to infer a structure, e.g., the score in CPN, from a known surface, e.g., a performance of a score, by application of Bayes' Rule, see Temperley (2007), i.e., if the probabilities of all possible performances of a score are known with regard to the probability of the score, and if the overall probabilities of the performances and the score are known, then the following relationship can be calculated:

$$P(score|performance) \propto P(performance|score)P(score) \tag{2.12}$$

The expression on the right side needs to be maximised in order to find the most likely score given a performance. In order to obtain the prior probability $P(score)$ all beat onset positions of the score on a normalised metrical grid are given statistical weights to appear at a specific onset time within the metrical grid. This requires the analysis of a known score. In order to achieve a broader application of the score prior, usually a corpus of scores from within the same style is used. This raises the question about the amount of learning that is actually required of the various algorithms in order to perform beat and tempo tracking reasonably well. This has obvious implications for the applications of beat tracking, for example score following. Composers often pursue the idea to break out of known styles, and how can the program make sure that a new piece is correctly beat tracked? Simply by knowing the score in advance, because it is necessary to train the algorithm about the probabilities of absolute beat positions. Another question that needs to be raised is the notion of tempo and beat that might not always be applicable to all possible musical scenarios. For example, the Italian *recitativo secco* does not have a beat structure but depends largely on the speech rhythm interrupted only by short cadential and thus metrical phrases of the *basso continuo* group of instruments. The question is

how a Bayesian beat tracker would handle the change of context in the above circumstance. It appears that such a beat tracker would only be as good as the prior distribution functions that have been chosen from specific musical styles and specific pieces that are felt to be representative or general enough in order for the algorithm to be successful. In the published material there is no indication how a beat tracker would handle the change of metre and tempo, for example a slow introduction to a movement is followed immediately by an Allegro in a different time signature. Beethoven's *Egmont Overture* has an Adagio introduction in 4/4 followed by a bridge at the end towards the Allegro in 3/4. Usually performances tend to accelerate the tempo in order to create a natural, dynamic transition and passage from the slow introduction to the fast main section of the piece. There also seems little consideration of the fermata problem within the literature. The question how a machine listening algorithm would handle musical passages that are 'out of time', for example the first sound of the *Egmont Overture* is held on a fermata. There are performers who strive to keep a logical connection between the length of a fermata sound and the following metrical structure, hence there might be a calculated, i.e., counted length of the fermata, but this is totally left to the discretion of the performers. Cemgil (2004) argues that the non-predictability of a performer's decision calls for a probabilistic approach. Nevertheless it seems hard to quantify the success of his proposed framework of algorithms simply because of the vast amount of musical realities in existence and the limited amount of published results that cover more than relatively easy musical examples drawn from folk or popular music with a steady beat, no sudden tempo and metre changes, no complex rhythms etc.. Unfortunately, Cemgil and Kappen have not published a detailed report on the numbers and kinds of errors generated by their system's output.

Raphael reports a probabilistic model for tempo detection combined with quantisation methods for transcription of performed note onsets. Raphael's method splits up into a rhythm process, a tempo process involving a random walk model and the observation data from the performance. MAP estimates are being solved by using dynamic programming. Of course as with other Bayesian techniques the system needs to be trained with multiple score performances in the same musical style before an actual performance is being analysed and transcribed by the system.

2.10.3 Pattern Matching

Kilian (2004), who works purely on MIDI data, uses a combination of pattern matching of groups of notes and statistical analysis for tempo detection evaluating single note durations. The pattern matching works on comparisons with a user-defined database of short rhythmic patterns. First, the MIDI performance data is merged into a single track of onset data with added information about the number of onsets at a given point, for example chords are detected and their number of notes kept together with their approximate mean onset time, similar to Dixon (2001), see also section 2.10.1 on page 69. The MIDI velocity of the notes played is also taken into account. The information so far collected informs a weight associated with each note/chord onset. Then, IOI ratios are calculated in order to detect durational accents,

i.e., changes between long and short notes and vice versa. Inverse durational accents (short-long syncopations) are filtered out from the list of onsets, which leads in effect to a rule-based low-pass filtering of IOI data. An explicit score database incorporates knowledge about rhythmic patterns in advance of the analysis. A distance function measures the distance between fragments of the performed rhythms and matched patterns of the database. The advantage of the pattern-matching approach is that the user can define the patterns within the database and expand or reduce its content. The disadvantage is that the database needs to cover all short patterns possibly occurring in the score that is being analysed. There is a manifold of patterns possible even if the method is restricted to patterns not exceeding the length of one bar, see for example figure 3-25 on Greek Verse Rhythms on page 103. Different musical styles or even differences within the same musical piece might turn the task of defining the best collection of patterns for the database to be difficult. Like in the Bayesian machine learning methods the user needs to enter information in advance of the process in order to allow the program to perform a detection task. In Kilian's approach patterns are allowed to overlap. A histogram-based approach is used for all onsets not explained by the pattern matching algorithm. Information about score durations mapped to previous IOIs influences the weight of a zone of attraction around simpler IOI ratios. In addition, an error detection penalises unusual score mappings within a given metrical hierarchy. Player characteristics can be taken into account by evaluating timing deviations from the IOI ratios $\frac{1}{1}$, $\frac{1}{2}$ and $\frac{2}{1}$.

2.11 Summary

We have described various forms of representations for musical rhythms and metre, for which CPN uses small-integer ratios in relation to a common reference duration 1/1, the semibreve. But there is, apart from the occasional use of metronome figures (BPM), no physical time incorporated in rhythm notation. In addition, apart from including words like *rubato* in the score, there exists no encoding of expressive timing deviations in the musical score, yet the flexible bending of performed durations is part of a truly musical performance, and it is perhaps one of the most important 'channels' for musicians to communicate with their audiences.

This means that an automated quantisation and transcription system has a major challenge, namely to arrive at a filtering method for expressive timing. It is crucial to be able to follow unpredictable tempo changes that have musical reasons from the point-of-view of the musician but it is difficult to detect them on the basis of the performance data only.

Therefore, there is a need to understand how various parts of the musical system of performance and perception works. It is crucial to look at the subject of temporal perception and we could learn about a range of perceptual timing thresholds. We have looked at musical metre and covered a seminal text by London (2004), in which he describes the perceptual and structural constraints for musical metre by introducing well-formedness constraints (WFC). He then also arrives at a *Many Meters Hypothesis* (MMH): Human listeners have learned throughout their life various tempo-modifications, rubati, and expressive timings, that are associated with a particular metrical pattern. This learning process is of course embedded within a specific

musical culture.

Musical ornaments, like trills and arpeggios, are challenging for an automated quantisation and transcription program. When executing ornaments, musicians will use a lot of freedom from metrical playing and even elements of improvisation, which can make their timing patterns unpredictable.

Throughout this work we will use the adjacent interval spectrum by Kjell Gustafson (Toussaint, 2004) in order to find a viable method for the detection of musical durations in spite of their constant variation in length. It was useful to learn about Gestalt principles, which has led us to develop a grouping algorithm that is used for the classification of performed durations, and which we are going to describe in chapter 5. For example, it is through the Gestalt principles of proximity and similarity within an adjacent interval spectrum of note durations that we are able to detect duration classes. The Gestalt principle of good continuation means that different duration classes can alternate in the course of a piece but it is still possible to differentiate between them and to detect reappearances of duration classes despite alternations and local tempo variations. Common fate is a Gestalt principle whereby groups of different duration classes are subject to the same tendency of expressive timing, i.e., a rallentando or accelerando will change the absolute time of every element of all duration classes in the same direction.

We have also discussed various methods of generating data sets, like onset detection and extraction from audio recordings, MIDI file analysis and the extraction of note-on events, real-time MIDI input and manual tapping using Csound. The noise levels introduced by these are minimal, and, when properly used, these methods stay below the levels where they would influence our quantisation program.

For the analysis of musical metre we have looked at various computational models of neural oscillations (Large and Kolen, 1995; Large, 2008; Large et al., 2010). This is a promising line of research and it supports our argument that human listeners approximate perceived rhythms and metrical patterns by tuning in on resonant oscillatory interactions of firing neurons, which have peak resonances at small-integer ratios.

We have also reviewed the current systems of quantisation and tempo tracking, which feature a wide range of different approaches. For quantisation, these include simple grid quantisers, context-free grammars, pattern-based approaches supported by user-created databases of rhythmic patterns, algorithmic approaches and Bayesian machine learning. Tempo tracking systems use clustering in combination with a multi-agent system, Bayesian machine learning techniques, pattern-matching, and computational models of neural oscillations, which resonate to rhythmic and metrical patterns of onsets.

Chapter 3

The Farey Sequence

3.1 Introduction

We have seen in section 2.2 that musical rhythm and metrical hierarchies are represented by small integer ratios in CPN. Note events fit into this framework by alignment with discrete onset times on a metrical grid structure, which itself represents a structured flow of continuous time. We will also learn that the order of mode locking regions in models of metre perception correlates with a tree of small integer ratios called the *Farey tree* (Large and Kolen, 1995). Also, one has found resonant frequencies in models of neural oscillators that show peak amplitudes at small integer ratios between the natural frequency of the oscillator and frequencies present in the rhythmic stimulus (Large et al., 2010). Therefore it seems justified to introduce the following unified model of musical rhythm at metre.

3.2 The Farey Sequence as a Model for Musical Rhythm and Metre

A Farey Sequence F_n can represent the onset times and relative durations of all metrical subdivisions of a beat. In this case, the onset times are marked by fractions between 0 and 1, which is the normalised duration of the beat.

As stated earlier, a fraction a/b belongs to F_n, if

$$0 <= a <= b <= n \tag{3.1}$$

with $a, b, n \in \mathbb{N}$. n is called the order of the Farey Sequence. Starting with $F_1 = \{\frac{0}{1}, \frac{1}{1}\}$, one can construct the next higher Farey Sequence by inserting mediant fractions between pairs of consecutive terms and by satisfying the above equation 3.1. A mediant fraction c/d between a/b and e/f has the form:

$$c/d = (a + e)/(b + f) \tag{3.2}$$

For example, to build F_2 one has to simply insert $\frac{1}{2}$ into F_1. The first few examples out of an infinite number of F_n are then as follows:

$$
\begin{aligned}
F_1 &= \{\tfrac{0}{1}, \tfrac{1}{1}\} \\
F_2 &= \{\tfrac{0}{1}, \tfrac{1}{2}, \tfrac{1}{1}\} \\
F_3 &= \{\tfrac{0}{1}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{1}{1}\} \\
F_4 &= \{\tfrac{0}{1}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{1}{1}\} \\
F_5 &= \{\tfrac{0}{1}, \tfrac{1}{5}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{2}{5}, \tfrac{1}{2}, \tfrac{3}{5}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{4}{5}, \tfrac{1}{1}\} \\
F_6 &= \{\tfrac{0}{1}, \tfrac{1}{6}, \tfrac{1}{5}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{2}{5}, \tfrac{1}{2}, \tfrac{3}{5}, \tfrac{2}{3}, \tfrac{3}{4}, \tfrac{4}{5}, \tfrac{5}{6}, \tfrac{1}{1}\}
\end{aligned}
$$

A Farey Sequence F_n has the following important properties:

1. If a/b and c/d are two successive terms of F_n, then $bc - ad = 1$.
2. If a/b, c/d, e/f are three successive terms of F_n, then $c/d = (a+e)/(b+f)$. c/d is called the mediant fraction between a/b and e/f.
3. If $n > 1$, then no two successive terms of F_n have the same denominator.

Next, we observe that F_n can represent all possible beat subdivisions. If a crotchet has the length 1, then the onsets of quaver subdivisions are the set $\{\frac{0}{1}, \frac{1}{2}\}$. The onsets of an eighth note triplet are the set $\{\frac{0}{1}, \frac{1}{3}, \frac{2}{3}\}$. Uniting both sets together with the set $\{\frac{1}{1}\}$ gives us F_3. The fraction $\frac{1}{1}$ marks the onset of the following crotchet. Rendering F_3 as a musical rhythm with two voices enables us to listen to the simultaneous rhythm of two quavers plus three events form a quaver triplet. Note how the Farey Sequence F_3 contains all onsets of this compound rhythm in order. Figure 3-1 shows the CPN for this example written as two voices. The durations of both voices have the ratio 2 : 3. Learning how to play this polyrhythm can be made easy by speaking the phrase: "Cold cup of tea". The third staff shows the notation of this rhythm as a compound polyrhythm. The next highest subdivision would be in four. The fractions to mark

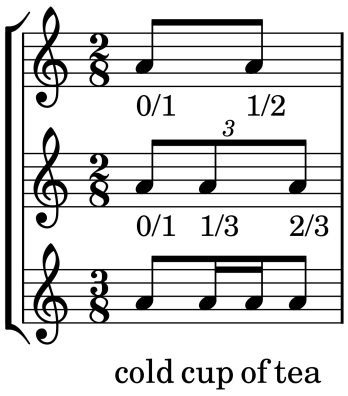


Figure 3-1: Rhythm of two quavers against three. Onsets marked by F_3. Its compound rhythm shown as the union of the two upper voices.

their onsets are part of F_4. But F_4 already comprises F_3 and F_2, which contain $\frac{1}{2}$. Therefore, and because all fractions must be in their lowest terms, only the set $\{\frac{1}{4}, \frac{3}{4}\}$ is added to F_3. This results in F_4 containing the onsets of the subdivisions $2:3:4$ of a crotchet with duration 1. One can continue to add higher subdivisions; F_n always contains the metrical subdivision n of a reference duration and all subdivisions m, where $1 < m <= n$, with $m, n \in \mathbb{N}$. In terms of CPN, F_n is a unified representation of the onsets of a polyrhythm. The integer ratios of the durations built from two sets of subdivisions are always $a:b$, with $0 < a < b <= n$.

The musical rendering of F_n as a compound polyrhythm reveals interesting structural properties. The IOIs between the onsets of the polyrhythms are the differences between consecutive pairs of fractions of F_n. Farey Sequences are symmetrical and self-similar sets of fractions. Rhythmically, the second half of the compound rhythm is the mirror of the first half. In musical terms, the onsets in the range of [0.5...1.0] form the *retrograde* of the first half of the Farey Sequence between 0 and 0.5. Retrograde pitch and rhythm structures have been used widely in Western music. They emerge from a set of transformations that is very common for polyphonic styles of music ranging from Renaissance to the Second Viennese School and beyond. A theme or a series of notes can be subjected to any of these transformations that comprise transposition, retrograde motion of pitch and rhythm structures, the inversion of pitch intervals and permutation of note pitches and durations in general, plus the retrograde of the inverted or permutated version of the theme or series of notes. The Farey Sequence comprises a rhythmic theme represented by the onsets between 0 and 0.5 *and* it continues with the retrograde version of the theme between 0.5 and 1. Of course, the retrograde version of the *entire* Farey Sequence would then produce the same rhythm as the forward version of it. The French composer Olivier Messiaen called this type of symmetric rhythm *non-retrogradable*[1] (Messiaen, 1995). Many examples of this technique can be found throughout his works, for example in the 5th movement of his composition *Visions de l'Amen*, see figure 3-2. This example is taken from Messiaen (1995, p.26). His non-retrogradable rhythms constitute a special form of subset

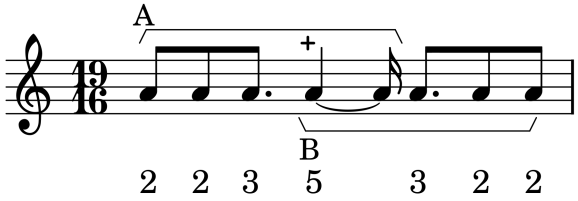


Figure 3-2: One of Messiaen's *non-retrogradable* rhythms. His technique uses a central duration, marked '+', around which a rhythmic pattern is mirrored: Pattern A is the mirror of pattern B. Durations are also shown in multiples of a semiquaver.

of a Farey Sequence, one that maintains mirror symmetry.

In order to notate a compound polyrhythm in CPN one would set all fractions in F_n to their lowest common denominator (lcd). Then, the sequence of their numerators denotes the

[1] In poetry a similar, non-reversable, mirror symmetric structure of letters, sounds or syllables is referred to as *Palindrome*, for example the famous "Sator Arepo Tenet Opera Rotas"

relative duration of every note of the polyrhythm. For its use in audio synthesis, one could simply multiply each numerator with the period assigned to the shortest duration. In order to model an non-isochronous metre for example, the period of pulsation of the N-cycle has to be >= 100 ms according to the WFC 1 (London, 2004), see also section 2.8.1.

Another question is how many distinct note onsets there are in a polyrhythm represented by F_n. This is easily answered by calculating the length of F_n. It depends on Euler's totient function $\varphi(m)$, which gives the number of integers <= m that are coprime to m:

$$|F_n| = 1 + \sum_{m=1}^{n} \varphi(m) \tag{3.3}$$

The totient function can be calculated using the following product of the primes dividing n:

$$\varphi(n) = n \cdot \prod_{p|n} (1 - \frac{1}{p}) \tag{3.4}$$

Table 3.1 shows the development of $|F_n|$ for the first few n. Note that where n is a prime number, $n - 1$ fractions are inserted into the previous Farey Sequence F_{n-1} .

n	$\lvert F_n \rvert$	n	$\lvert F_n \rvert$	n	$\lvert F_n \rvert$
1	2	11	43	30	279
2	3	12	47	40	491
3	5	13	59	50	775
4	7	14	65	60	1103
5	11	15	73	70	1495
6	13	16	81	80	1967
7	19	17	97	90	2481
8	23	18	103	100	3045
9	29	19	121	200	12233
10	33	20	129		

Table 3.1: Development of the length of F_n for some values of n.

Rhythms consist usually of a chain of instances of specific note durations. We use the word 'chain' in the metaphorical sense. We imagine onsets of rhythms as points linked together on a timeline and the duration between a pair of onsets, the IOI, constitutes the link between them. A metre consists either of isochronous or non-isochronous beat cycles. Metres are cyclic chains of instances of a beat with all its possible subdivisions. Similar to these examples, one can link instances of Farey Sequences together, one after the other, to form a new, single and larger sequence. And by doing so, one would demand that the overall range of fractions remains between 0 and 1, so that the resulting sequence remains an ordered set of unique fractions - a necessary condition to represent chains of onsets on a continuous timeline. Would the resulting sequence still satisfy the definition of a Farey Sequence? Clearly not, because one or more subdivisions contained in the higher Farey Sequence will be missing. It turns out that the sequencing of smaller F_n, as described above, always creates a subset of a larger F_k. First,

consider the case of two identical Farey Sequences F_n linked together. Fractions of both F_n range between 0 and 1. In order to form the subset of a larger Farey Sequence, F_{n+n}, the range of the participating sequences needs to be scaled; the first sequence into the range between 0 and 0.5, the second into the range between 0.5 and 1. Because two Farey Sequences are involved, the scaling factor is 2. Using two instances of F_4 as an example, here is what needs to be done in order to transform them into a subset of F_8: First, the two sets of fractions have to be scaled by two and the resulting fractions reduced to their lowest terms. We introduce the notation $F_n[x, y]$ with $x, y \in \mathbb{R}$ to indicate the scaling of the fractions in F_n into the range given by the two real numbers x, y.

$$\{F_4, F_4\} = \{\{\frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}\}, \{\frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}\}\}$$

$$\{F_4[0, 0.5], F_4[0.5, 1]\} = \{\{\frac{0}{1}, \frac{1}{8}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{3}{8}, \frac{1}{2}\}, \{\frac{1}{2}, \frac{5}{8}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}, \frac{7}{8}, \frac{1}{1}\}\}$$

Then, by process of unification, the duplicate term $\frac{1}{2}$ is deleted, resulting in an ordered subset of F_8:

$$F_4[0, 0.5] \cup F_4[0.5, 1] = \{\frac{0}{1}, \frac{1}{8}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6}, \frac{7}{8}, \frac{1}{1}\} \subset F_8$$

$$F_8 = \{\frac{0}{1}, \frac{1}{8}, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{3}{8}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{5}{8}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \frac{7}{8}, \frac{1}{1}\}$$

It is evident that the onset times of the initial sequences change, but the ratios of the durations (IOIs) brought into the resulting subset remain the same, due the linear scaling. Forming a subset like the above is equivalent with taking a higher order Farey Sequence and filtering out those fractions, which have denominators composed of higher prime numbers, 5 and 7 in the above example. The principle of composing longer sequences also works when the order n of the Farey Sequences involved is not equal. We note that a subset of F_n, which is obtained by sequencing, is a filtered Farey Sequence, where the filtering method will involve the prime number composition of the denominators of the fractions in F_n. The filtering and the concatenation of Farey Sequences can be used both in order to represent musical rhythm and metre. We found that a filtered Farey Sequence can contain linearly scaled versions of complete Farey Sequences of lower order. This is possible because the Farey Sequence has a self-similar structure. Finally, if F_n can model musical metre and rhythms within a single bar, then sequenced instances of F_n are able to model an entire score.

3.2.1 Building Consecutive Ratios Anywhere in Farey Sequences

Given that the distance between the elements of a Farey Sequence is always $c/d - a/b = 1/db$ and if we only know the fraction a/b given in its lowest terms, i.e., a and b are coprime, then building the harmonic series:

$$\{\ldots, -1/3b, -1/2b, -1/b, 1/b, 1/2b, 1/3b, \ldots\}$$

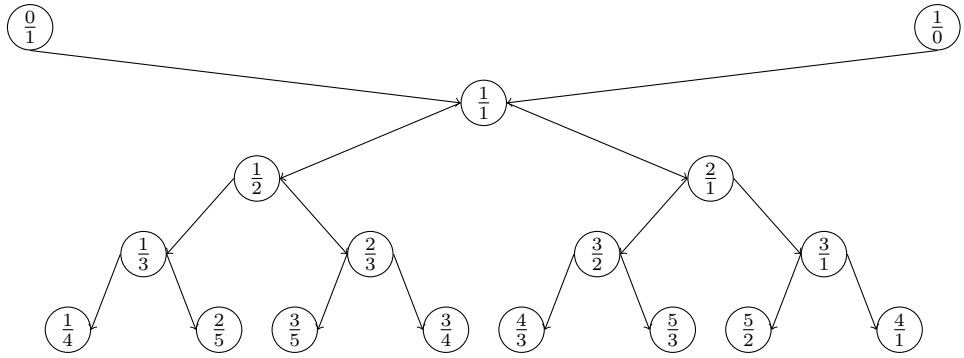


Figure 3-3: The Stern-Brocot tree. Its left-hand branch growing from 0/1 and 1/1 is called the Farey tree.

gives all possible distances $|1/kb|$ from a/b with $k \in \mathbb{N}$ where all fractions ever adjacent to $a/b \in F_n$ could be found, no matter how high n becomes. Of course, not all coefficients k lead to adjacent fractions. The search for the fraction c/d adjacent to a/b has to be constrained by the relation: $ad - bc = 1$ if $c/d < a/b$. Similarly, $bc - ad = 1$ if $c/d > a/b$. If we furthermore choose k so that d becomes the largest value $<= n$ fulfilling that relationship, then we always find the preceding and the succeeding ratio adjacent to $a/b \in F_n$. This has already been proven by Hardy and Wright (1938, pp.25). We have therefore implemented an algorithm that is able to construct subsets of F_n of any length $1 <= m <= |F_n|$ given only one member of F_n as a seed value. This is useful because for large n the construction of the *entire* list of ratios $\in F_n$ can take a considerable amount of computing time and memory space, because the length of F_n grows rapidly with n, see equation 3.3 and table 3.1.

3.2.2 The Farey Sequence, Arnol'd Tongues and the Stern-Brocot tree

The Farey Sequence is a subtree of the Stern-Brocot tree (Graham et al., 1994; Calkin and Wilf, 2000; Sloane, 2011). The Stern-Brocot tree is generated by starting with the fractions 0/1 and 1/0. In each iteration one inserts a mediant fraction between each pair of fractions obtained from the previous iteration. Note that this generating procedure is different from the Farey Sequence, because the size of the denominator is not constrained as in equation 3.1.

Starting with the root 1/1 and per subsequent iteration, each new mediant fraction forms a new node, which is then connected to the nearest fraction or node obtained from the previous iteration. This builds the following binary tree, see figure 3-3. Note that the left-hand side of the tree, going down from 1/2, yields fractions, which are elements of Farey Sequences. But note also that on the same level of the tree there are fractions which may belong to two (or more) different Farey Sequences, for example the fourth iteration adds the fractions $\frac{1}{4}, \frac{2}{5}, \frac{3}{5}, \frac{3}{4}$. Although $\frac{1}{4}$ and $\frac{3}{4}$ complete Farey Sequence F_4, $\frac{2}{5}$ and $\frac{3}{5}$ are only two of the five fractions that would be needed to obtain F_5. We note that the nth iteration for the Stern-Brocot tree does not

generate the Farey Sequence F_n on the left-hand branch of the tree. It only generates a subset. The literature refers to this branch as *Farey tree* (Large and Kolen, 1995; Schroeder, 1991, p.336). Here the integer ratios of a Farey tree mark the strength and ordering of mode locking regions, also known as Arnol'd Tongues, in a dynamic system of weakly coupled oscillators.

3.2.3 Farey Sequences and Musical Rhythms

We have seen that it is possible to use the Farey Sequence as the principal form of representation of rhythms in CPN. We will show that it can also be used in order to represent plausible approximations of non-notated musical forms, such as jazz improvisation, world music, etc.. This method uses Farey Sequences to form grid points for rhythm quantisation; the outcome of the quantisation process is then a sub-set of some Farey Sequence F_n. We showed already in Boenn (2007b) that Farey Sequences can represent rhythms of many diverse styles from different historical and cultural contexts, because of their scalability and self-similarity and because appropriate filters can be found. Results obtained by Desain and Honing (2003) and Papadelis and Papanikolaou (2004) suggest that human listeners categorise performed rhythms in such a way that classes of short rhythmic patterns are formed in the vicinity of relatively small integer ratios. Again, there are filtered Farey Sequences capable of representing such categories.

Graham et al. (1994) have shown that the Stern-Brocot tree can serve to approximate decimal fractions by integer fractions, with arbitrary precision. See also Martelli et al. (2005, pp.675) for a Python algorithm that approximates floating point numbers into Farey fractions. There are also compositional applications that can be derived from Farey Sequences as shown in Boenn (2007b). One of the main advantages of the Farey Sequence is that it is scalable to different musical and cognitive timescales and does not rely on context-related manifestations of metre and bar. However, metrical hierarchies on any level can easily be modelled using filtered Farey Sequences.

Musically Relevant Structure of the Farey Sequence Figure 3-4 shows a plot of F_{17} that correlates the position of the ratios a/b in the interval $]0 \ldots 1[$ with the unit fraction $1/b$ built from their corresponding denominators. The unit fractions represent the metrical subdivision by b of the reference duration 1. The plot illustrates quite nicely the self-similarity of a Farey Sequence. The visualisation also makes clear that the smaller the denominator, the larger are the symmetrical gaps around the x-position of the ratio, i.e., the smaller b of the ratio a/b, the greater the distance to the next ratio with the denominator $b + 1$. We believe that it is due to those relatively large gaps surrounding small integer ratios that they form a zone of higher probability for rhythm and metre perception to lock onto them. Onset time ratios that fall into these zones are shifted towards the ratio that exhibits a local peak of attentional energy (Large and Palmer, 2002). Another way of looking at the space immediately around small integer ratios is to note that the denominators in the immediate vicinity of those ratios have the tendency to become larger and larger, in fact when n goes to infinity, then the denominators of the ratios,

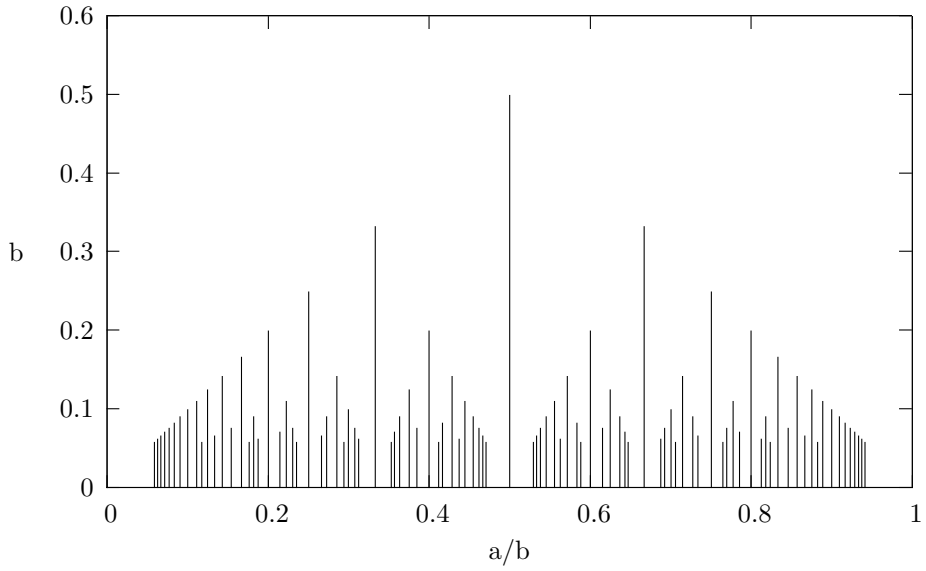


Figure 3-4: Correlation of $a/b \in F_{17}$ and $1/b$ in the interval $]0 \ldots 1[$

who mark their positions on the x-axis close to small integer ratios, also go to infinity. One can see this phenomenon indicated in figure 3-4. Around the value $x = 0.5 = \frac{1}{2}$ there is a tendency for denominators of surrounding fractions to become gradually larger and larger as one approaches $x = 0.5$ using fractions $\frac{a}{b}$ close to $\frac{1}{2}$, i.e., the value of $\frac{1}{b}$ becomes smaller and smaller. Can it be that this value of $\frac{1}{b}$ at normalised time positions given by $\frac{a}{b}$ is perceptually relevant? There is experimental evidence that zones of higher perceptual energy around small integer ratios do indeed exist (Large and Palmer, 2002). This will be an important result to consider in the design of software for quantisation and tempo tracking of musically performed events. Composers who want to "stay away" from those simple ratios will need to leave a considerable amount of space around these zones of attraction, or they would need to blur the emergence of small integer ratios by using a dust of events around time points of high metrical expectancy. This technique can be seen in many works of 20th century avantgarde music, for example in the scores of Ferneyhough, Nono, or the school of "new complexity", who use much higher degrees of metrical subdivision, often also in nested combinations, as compared to earlier examples of music history. One can also expect that the onset times from musical performances of rhythms that are aligned to a metrical grid exhibit characteristic deviations from the small integer ratios used for CPN. Above we have seen that only small deviations produce ratios with relatively high denominators and with relatively high prime numbers as their components. We identify this as the main problem for rhythm transcription from performance data. The inter-onset ratios recorded from performances usually need to undergo a small shift towards a ratio that is normally used for CPN. Such ratios usually only involve small prime number divisors of their denominators, so-called k-smooth numbers (Berndt, 1994; Blecksmith et al., 1998).

To continue our analysis, there are various *accelerandi* and *ritardandi* encoded in every F_n, for example there are Gestalten that form visible triangles between larger reciprocals and smaller ones in their surrounding area, see 3-4. These Gestalten are formed by monotonically increasing or decreasing values of the denominators. The increasing or decreasing tendencies can overlap each other. The gaps between the ratios forming those triangles are always on a logarithmic scale, hence the impression of accelerated or reduced tempo that becomes evident through the sonification of these triangular Getstalten. It is well known that timed durations need to be placed on a logarithmic rather than linear scale in order to convey a "natural" sense of spacing and tempo modification (Essl, 1996). These structures are clearly mirrored around the 1/2 value that is part of every F_n with $n > 1$.

If we take the space between 0 and 1 as the normalised duration of a musical sequence with a period of x milliseconds, then there is only a certain degree of subdivision perceptually viable as an element of a rhythmic figure within that sequence (London, 2004). If we go a step further and investigate also compound rhythms emerging between various levels of subdivisions then we must take into account the subdivision of the beat based on the least common denominator between those two levels of subdivision of the beat that are involved in producing the shortest inter-onset intervals. The subdivisions responsible for producing the shortest IOI between their elements are easy to find for a Farey Sequence. For any F_n, the shortest IOI will be given by $\frac{1}{n} - \frac{1}{n-1}$. In order to apply London's constraint one has to filter out certain subdivisions according to a specific minimum tempo period x.

3.2.4 The Farey Sequence and Musical Notation

Our motivation is to provide a general representation of musical rhythm and metre and to allow for the quantisation of performed musical rhythms into CPN. This representation shall be modelled through the use of Farey Sequences. In order for this to work we need to investigate whether it is possible to represent any score written in CPN as Farey Sequences and that it is possible to go back and forth between the two. Therefore, the aim is to find a model general enough in order to represent rhythms in CPN, but which is also capable of dealing with rhythms and metres of non-Western musical cultures as long as they are based on an underlying pulsation. This is the purpose of this subsection. Our hypothesis is therefore that every rhythm in pulsed music can be represented precisely by a complete or filtered Farey sequence, as long as the rhythm is written in CPN or another form of proportional representation.

This is the general method: Take the total length of the notated rhythm in beats and multiply by the ratio of the shortest rhythmic value to the beat. This gives you the index n of the appropriate Farey sequence F_n. Then, note duration values of the rhythm can be mapped precisely to F_n, thus representing the ratios of onset times, i.e., the occurrence of note events on the timeline.

Figure 3-5, for example, shows a rhythm with six beats. The beat pulses in crotchets, the shortest value in the sequence is a quaver. Thus, their ratio is $\frac{2}{1}$, and the Farey sequence index

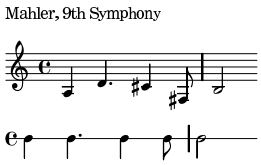


Figure 3-5: Rhythm of Horn Motive from the 1st Movement of Mahler's *9th Symphony.*

is

$$n = 6 \times \frac{2}{1} = 12$$

The Farey sequence with smallest n that contains the rhythm is therefore F_{12}, representing minims as $\frac{1}{3}$, crotchets as $\frac{1}{6}$, and quavers as $\frac{1}{12}$. Every rhythm of six beats whose shortest value is a quaver can be found within F_{12}.

The occurrences of the onsets of the Mahler rhythm in figure 3-5 within F_{12} are

$$\frac{0}{1}, \frac{1}{6}, \frac{5}{12}, \frac{7}{12}, \frac{2}{3}.$$

The durations are derived from here as the following inter-onset time ratios r_{iot}:

$$r_{iot} = \{\frac{2}{12}, \frac{3}{12}, \frac{2}{12}, \frac{1}{12}, \frac{4}{12}\}.$$

There are two ways of translating rhythms from their representation as a filtered Farey sequence back into Western music notation:

1. the multiplicative way,

2. by making subdivisions.

The first method needs to transform the list of inter-onset time ratios into a list of ratios with the lowest common denominator. From there we need to know the shortest note value, which is the smallest pulse duration p_d, in our example it is $\frac{1}{12}$.

The graphic result also depends on how that pulse duration is chosen to be represented. We call this the pulse representation p_r, e.g., $\frac{1}{12}$ is regarded as a quaver ($\frac{1}{8}$) within CPN. The knowledge of those values leads to a notation where the notes are being represented by integer multiples of the pulse representation.

Given our Mahler example above, the set of note durations r_{glyph} is calculated with the formula

$$r_{glyph} = \{\frac{p_r x}{p_d} | x \in r_{iot}\},$$

which after reducing into lowest terms leads to

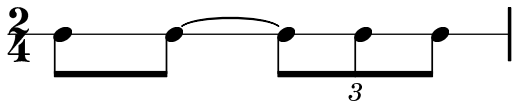
$$r_{glyph} = \{\frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{8}, \frac{1}{2}\}.$$

As you can see, all values of this list can be mapped directly to a symbol in CPN except $\frac{3}{8}$, which will be recognised as a dotted note value: $\frac{1}{4} + \frac{1}{8}$.

This example has shown that it is possible to convert rhythms in CPN into a subset of F_n and back again into symbols for CPN.

The next example will use compound note values in CPN and demonstrate how to convert the normalised inter-onset ratios back into CPN. Take the following rhythm within F_6:

$$\frac{0}{1}, \frac{1}{4}, \frac{2}{3}, \frac{5}{6}.$$



This is a rhythm that contains a tie between a quaver and a triplet note. The inter-onset ratios are

$$r_{iot} = \{\frac{1}{4}, \frac{5}{12}, \frac{1}{6}, \frac{1}{6}\}.$$

In order to translate those ratios into note values, one needs to scale them according to the reference value of the notation, which we call the beat representation b. In this case we choose b to be a crotchet: $b = \frac{1}{4}$.

We then choose a value for the number of beats n in which we would like our rhythm to be fit, e.g., $n = 2$ yields a notation that represents our example within a $\frac{2}{4}$ bar. For a list of note values we calculate the total length $L = \frac{1}{2}$ of the rhythm as a note value according to:

$$L = b \times n$$

We are then to multiply each item in the list of inter-onset ratios by L and we obtain the list of note values we are looking for.

$$r_{glyph} = \{Lx | x \in r_{iot}\}$$

resulting in

$$r_{glyph} = \{\frac{1}{8}, \frac{5}{24}, \frac{1}{12}, \frac{1}{12}\}.$$

Again we can see that every unit fraction in this list can directly be written as a symbol in CPN. Therefore, all non-unit fractions have to be further processed.

Since the fraction $\frac{5}{24}$ irreducible it needs to be written as a sum of fractions with equal denominator. Therefore we calculate all partitions of the numerator and we choose a partition that meets the following two requirements:

1. The terms of the partition must have common divisors with the denominator, i.e., they must not be coprime, and

2. the partition with the lowest number of terms is preferred.

5
$4 + 1$
$3 + 2$
$3 + 1 + 1$
$2 + 2 + 1$
$2 + 1 + 1 + 1$
$1 + 1 + 1 + 1 + 1$

Table 3.2: All partitions of 5.

In the r_{glyph} of our example, $\frac{5}{24}$ is irreducible. A list of all partitions of 5 can be seen in table 3.2. Two partitions ($4 + 1$ and $3 + 2$) meet the requirements, which indicates also that there must be a tie between two note values or perhaps a dotted note in the score.

We then proceed with building the fractions and write them in lowest terms

$$s_1 = \frac{5}{24} = \frac{4}{24} + \frac{1}{24} = \frac{1}{6} + \frac{1}{24}$$

and

$$s_2 = \frac{5}{24} = \frac{3}{24} + \frac{2}{24} = \frac{1}{8} + \frac{1}{12}.$$

Both results s_1 and s_2 are not equivalent with respect to the standards of CPN. In order to make a clear decision between the two possibilities we again find the Farey Sequence very useful. When we try to fit both results into the rhythm using the Farey Sequence F_6 from where we started, we will observe some beneficial differences between them.

First we need to reverse the scaling by L, then take the inter-onset ratio $\frac{1}{4}$, which indicates the point when the new note value should start, build the sums and compare the results with the members of F_6.

We start with scaling s_1:

$$s_1 = \{\frac{1}{6} \times 2 = \frac{1}{3}, \frac{1}{24} \times 2 = \frac{1}{12}\}$$

Then we check the sums with $\frac{1}{4}$ to see whether they are members of F_6. We see that that the sum

$$\frac{1}{4} + \frac{1}{3} = \frac{7}{12}$$

fails that test.

We do the same test with the second member of s_1:

$$\frac{1}{4} + \frac{1}{12} = \frac{1}{3}$$

This time the test succeeds, $\frac{1}{3}$ is indeed a member of F_6.

When we further build on this fraction and do the second sum

$$\frac{1}{3} + \frac{1}{3} = \frac{2}{3},$$

s_1 seems to fit nicely, however $\frac{1}{3}$ and $\frac{2}{3}$ are both members of F_6 but neither of them lines up within the *inter-beat ratios*

$$r_{ib} = \{\frac{0}{1}, \frac{1}{2}, \frac{1}{1}\}.$$

The explanation is simple if one thinks about the tie in our example. The tie is necessary because there is a transition between an eighth pulsation and an eighth triplet pulsation. Those kind of transitions only take place when the borderline to the next beat is crossed.

Therefore we need to consider those borders as they materialise exactly when a new beat occurs. Our example contains three such borderlines (see r_{ib}). If one would erase the tie, the second borderline would materialise itself at $\frac{1}{2}$.

Now look at s_2. After scaling

$$s_2 = \{\frac{1}{8} \times 2 = \frac{1}{4}, \frac{1}{12} \times 2 = \frac{1}{6}\},$$

we check the sums with $\frac{1}{4}$.

$$\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

This is both part of F_6 and a match with the inter-beat ratio $\frac{1}{2}$.

Carrying on with $\frac{1}{2}$ yields

$$\frac{1}{2} + \frac{1}{6} = \frac{2}{3},$$

which is contained within F_6 as well, so s_2 appears to be the solution for resolving $\frac{5}{24}$ into notation glyphs.

Let us do a reverse order in s_2 and write

$$\frac{1}{4} + \frac{1}{6} = \frac{5}{12}.$$

This test fails because $\frac{5}{12}$ is not in F_6.

If we would continue with $\frac{5}{12}$ we would match up with F_6 again:

$$\frac{5}{12} + \frac{1}{4} = \frac{2}{3}.$$

Apparently, the order of the ratios is crucial to our solution and it can be tested by using the smallest F_n that forms the grid for the rhythm and a test for matching inter-beat ratios who are automatically members of the F_n sequence as well. From our example we conclude that at least one of the ratios has to meet the last requirement.

We now have sufficient tools together to transcribe a sequence of inter-onset ratios into a sequence of ratios that can directly be used for music notation. If any ratio is encountered that has coprime components, then the ratio needs to be written as a sum of smaller ratios that can be found by making partitions of the nominator and choose the right pattern according to the three following rules:

1. All terms of the partition must have common divisors with the denominator, i.e., they must not be coprime,

2. the partition with the lowest number of terms is preferred, and

3. at least one of the resulting ratios has to match an inter-beat ratio.

In addition, if two of the resulting ratios itself, next to each other, have the extra property of being at a ratio of $2:1$, then the result is a *dotted* note value,

e.g., $\frac{3}{8} = \frac{1}{4} + \frac{1}{8}$ will be written as a dotted crotchet. If we encounter a sequence of three ratios that in sequence are at a ratio of $1:2, 1:2$, the result will be a double dotted note value, e.g., $\frac{7}{16} = \frac{1}{4} + \frac{1}{8} + \frac{1}{16}$

Conclusion: Every Farey sequence F_n can be regarded rhythmically as a superposition of all pulsations with inter-onset ratios ranging between $\frac{1}{1}$ and $\frac{1}{n}$ and therefore there exists a F_n for every rhythm in pulsed music with rational inter-onset ratios. For every rhythm written in CPN there is a F_n, or at least one subset, that can represent this rhythm. The transformation of note onsets and durations written in CPN into elements of F_n can also be reversed. This means that composers can exploit the Farey Sequence for algorithmic composition and convert every rhythm conceived on this basis into CPN.

3.3 Filtered Farey Sequences

3.3.1 Introduction

A Farey Sequence can represent the onset times and relative durations of metrical subdivisions of a beat. The concatenation of various F_n into a sequence forms subsets of a higher order F_{n+k}. Therefore, appropriate filtering methods can model any musical rhythm as a filtered instance of F_{n+k}. Such a filtered sequence that represents onsets and IOIs of musical events can model the compound rhythm of any musical sequence that is either performed, read from a score or algorithmically generated by a computer. The grid of timestamps underlying the recorded onsets should have a period $T \leq 2ms$. Normalised onset times can be converted into integer ratios via continued fraction expansion. A maximum of ten recursions within the expansion algorithm provides sufficient precision. The highest denominator in a sequence of such onsets determines the order n of the Farey Sequence F_n that forms the superset for all onset fractions.

The compound rhythms and normalised onset times of a musical performance form a subset $x \subset F_n$. F_n will contain another subset y that holds the onset times of the score representation of the performed rhythm. We expect x to have a large amount of elements with relatively large integers as denominators, whereas the set of score onsets y features mostly fractions with relatively small denominators. The highest prime number component of all metrical subdivisions within the score indicates the smoothness of the denominators found in y. A positive integer is k-smooth if its prime divisors are $\leq k$ (Berndt, 1994; Blecksmith et al., 1998). In our case we are interested in the smoothness of the denominator of a fraction in y.

For example, if the highest prime number subdivision of all note durations within the score is 3, then all the denominators of the fractions that represent note onsets must belong to the *3-smooth* numbers, which means they can only be products of the powers of 2 and 3. The performed onsets in x introduce deviations from the score onsets because of expressive timing, therefore one can expect that the fractions in x will have relatively large denominators with prime divisors > 3. Only a machine performance of the score onsets y would return an identical set $x = y$. Therefore it is appropriate to look at x and y as subsets of F_n with a high probability of the relation $x \neq y$. In our previous analysis we have seen that the subsets x and y are both filtered Farey Sequences or scaled versions of a complete Farey Sequence of lower order. The question is whether there can be an algorithm that can analyse the sets x and y in such a way that it is able to map each member of x uniquely to a member in y. This would allow us to use a Farey Sequence not only to represent the onset times of an entire score, but also to find a mapping of one subset of a Farey Sequence that represents performed onset times to a subset of F_n, which represents the score onset times of that performance. We will discuss in chapter 6 how this can be achieved.

3.3.2 Polyrhythms and Polyphony

Hemiola A very common case of polyrhythm in Western music is the so-called *hemiola*. This two-voice rhythm has a proportion of 2 : 3 and can be modelled with F_3, as shown in figure 3-6 with the compound rhythm and metre added as an extra voice. This kind of hemiola

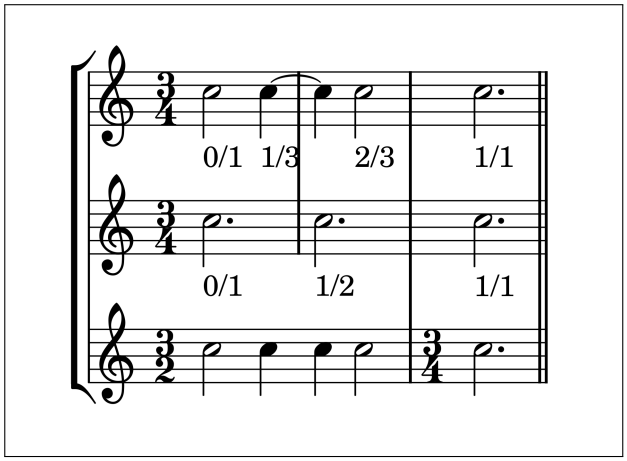


Figure 3-6: Hemiola with onsets marked by F_3

can be heard very often in ornamented cadences of the Baroque era, see figures 3-7 and 3-8.

Throughout history, one finds many musical examples where composers make use of this rhythmic device. In figure 3-9 we see the beginning of Robert Schumann's 3rd Symphony in E-flat major. The hemiola here is insofar unusual as it sits right at the opening of the first movement with its principle theme. The actual 3/4 metre of the piece is only established at the arrival of the 7th bar.

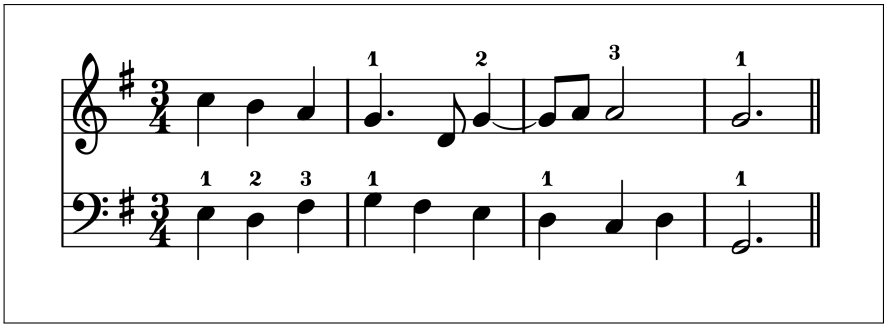


Figure 3-7: A typical Baroque cadence using a hemiola.

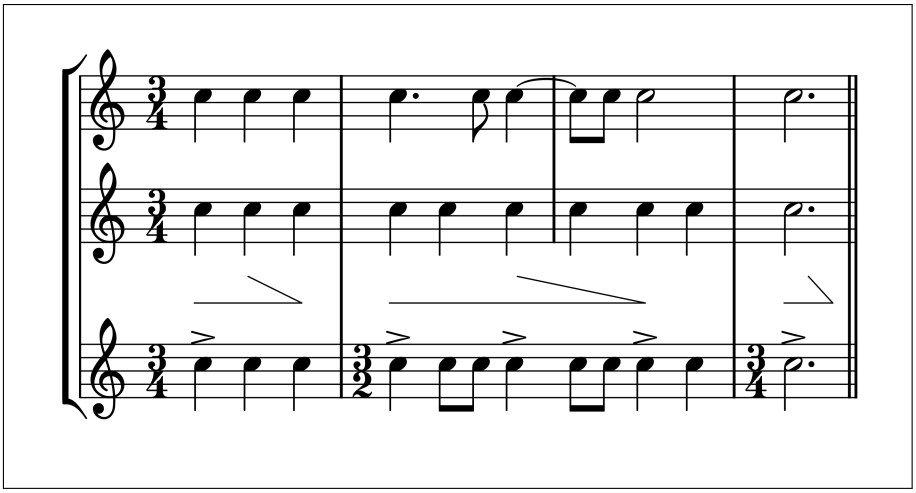


Figure 3-8: How the compound rhythm of the hemiola in figure 3-7 can be interpreted as a 3/2 bar. The open triangle symbols above the third staff show how a conductor would possibly beat this passage.

Polyrhythms as Compound Rhythms F_n can also represent polyrhythms of the highest complexity whilst being independent from graphical score representation. Figure 3-10 shows the compound rhythm representation of the polyrhythm 6:5:4:3:2:1 as encoded by equation 3.5.

$$F_6 = \{\frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1}\} \tag{3.5}$$

Although figure 3-11 notates the same rhythm as a score with five voices, note the differences between this polyphonic voice notation and our representation as a monophonic compound rhythm, where its main advantage is to display the IOIs between every single event of the polyrhythm. It is also an example to show how our transcription algorithm (described in section 6.2.1) is able to find the least complex score representation. The notation generated by our tool enables a solo instrument to learn to play complex polyrhythms in a much easier way. This outcome is also helpful when writing for larger ensembles or orchestras because of their demand for a clear notation that helps players to learn a new piece in a short period of time.

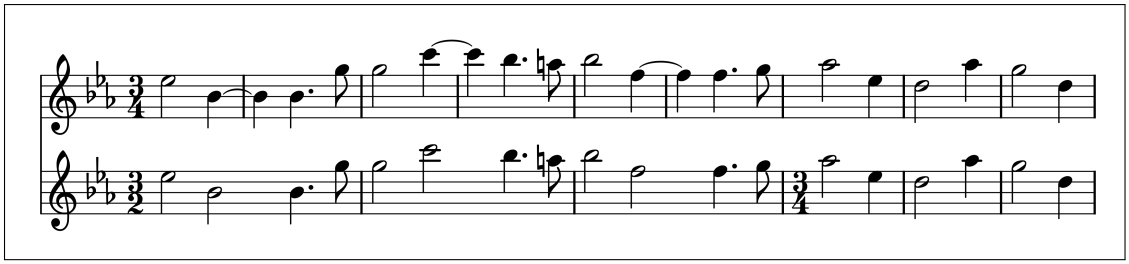


Figure 3-9: Schumann's Third Symphony begins with the main theme in three hemiolas before picking up the 3/4 metre of the movement. The entire orchestra plays the same rhythm during the first six bars. The lower staff shows how the hemiolas are musically understood.

I	$\frac{0}{1}$						$\frac{1}{6}$						$\frac{1}{3}$					
II			$\frac{1}{18}$		$\frac{1}{9}$			$\frac{2}{9}$		$\frac{5}{18}$				$\frac{7}{18}$		$\frac{4}{9}$		
III		$\frac{1}{36}$		$\frac{1}{12}$		$\frac{5}{36}$		$\frac{7}{36}$		$\frac{1}{4}$		$\frac{11}{36}$		$\frac{13}{36}$		$\frac{5}{12}$		$\frac{17}{36}$
I	$\frac{1}{2}$						$\frac{2}{3}$						$\frac{5}{6}$					
II			$\frac{5}{9}$		$\frac{11}{18}$			$\frac{13}{18}$		$\frac{7}{9}$				$\frac{8}{9}$		$\frac{17}{18}$		
III		$\frac{19}{36}$		$\frac{7}{12}$		$\frac{23}{36}$		$\frac{25}{36}$		$\frac{3}{4}$		$\frac{29}{36}$		$\frac{31}{36}$		$\frac{11}{12}$		$\frac{35}{36}$

Table 3.3: Onset times of the metrical subdivisions underlying the first voice in figure 3-12 expressed as F_{36}. The structure of subdivision is ◯, *tempus perfectum cum prolatione imperfecta*. This meta-cycle encompasses 6 mensurations of 3 x 2 minims at the beginning of Ockeghem's Credo. The beat period is $\frac{1}{36}$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
D	X	o	O	o	O	o	X	o	O	o	O	o	X	o	O	o	O	o
A	X	o	O	o	X	o	O	o	X	o	O	o	X	o	O	o	X	o
T	X	o	o	O	o	o	O	o	o	X	o	o	O	o	o	O	o	o
B	X	o	o	O	o	o	X	o	o	O	o	o	X	o	o	O	o	o
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
D	X	o	O	o	O	o	X	o	O	o	O	o	X	o	O	o	O	o
A	O	o	X	o	O	o	X	o	O	o	X	o	O	o	X	o	O	o
T	X	o	o	O	o	o	O	o	o	X	o	o	O	o	o	O	o	o
B	X	o	o	O	o	o	X	o	o	O	o	o	X	o	o	O	o	o

Table 3.4: Pseudo-Polymetric structure of beats from Ockeghem's Missa prolationem creating a meta-cycle of 36 beats. 'X' marks the beginning of a new mensuration (equivalent to today's downbeat), 'O' marks a 2nd level beat, 'o' marks a 3rd level beat.

Ockeghem Polyrhythmic concepts have been revisited by 20th century avantgarde composers who, like Olivier Messiaen or Steve Reich, adapted the concept from non-Western musical cultures. Others may have been inspired by 15th century composer Johannes Ockeghem who composed his *Missa prolationum* using a sophisticated structure of two simultaneous double-canons, whereby, for each canon, two voices use the same pitch sequences and rhythmic values but each of the voices execute them using their own tempo. All voices do this simultaneously, thereby creating a complex vocal polyphony, see figure 3-12.

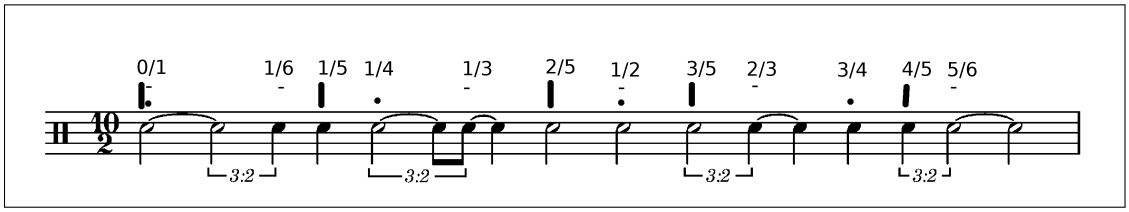


Figure 3-10: The compound polyrhythm of F_6. The entire bar represents the space between $\frac{0}{1}$ and $\frac{1}{1}$. '-' marks the onsets of subdivision in 6, '|' marks subdivision in 5, '.' marks subdivision in 4. See also figure 3-11.

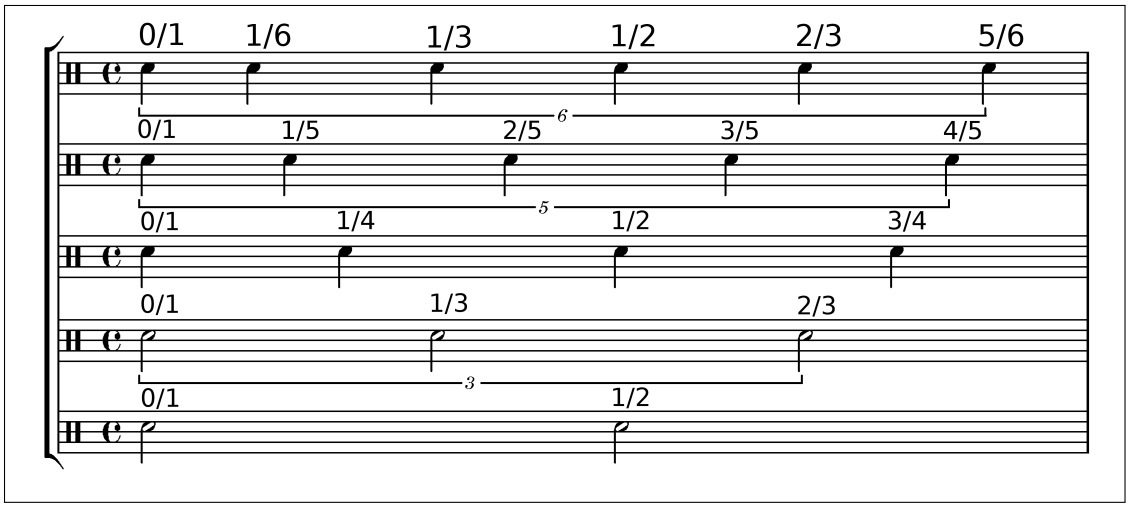


Figure 3-11: The polyrhythm of F_6 split into 5 voices. The entire bar represents the space between $\frac{0}{1}$ and $\frac{1}{1}$. Each subdivision is represented by its own voice. If one would merge all five voices into a single line, then figure 3-10 shows the least complex notation for this problem.

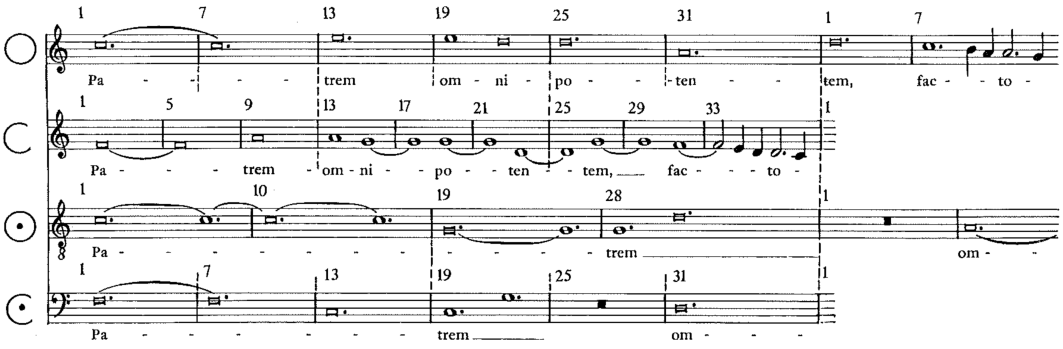


Figure 3-12: The beginning of the Credo of from the *Missa Prolationum* transcribed into CPN. Beats are numbered 1-36, 'downbeats' are marked with a beat number. The circular signs at the beginning indicate four different 'tempi', i.e., schemes of metrical subdivision (de la Motte, 1981). Reproduced with permission. See also figure 3-13.

Creation of four different tempi using the same sets of notes and rhythms was possible because the duration of the brevis, which translates into today's double whole-note, was not fixed as in CPN today. There were many different options for the composer of that era, as for example outlined in the famous 14th century treatise 'Ars Nova' by Philippe de Vitry (de Vitry, 1964). The brevis can be interpreted either as 3 or 2 semibreve, and those further into 3 and 2 subdivisions called minims, so we can subdivide the brevis into 3 x 2, 2 x 2, 3 x 3 or 2 x 3 minims. This was indicated in the vocal score via different mensuration signs, from which today's time-signatures devloped, see figure 3-13. Ockeghem uses all four possibilities simultaneously, thus introducing a proportion of 6:4:9:6 between the voices. The effect of superimposing these different subdivisions can be seen in Table 3.4. When transcribed nowadays into CPN the voices would need 36 minims to complete the full cycle of patterns and start on the same 'downbeat' again (de la Motte, 1981). The complete cycle translates into a *filtered* Farey Sequence F_{36}. This demonstrates the ingenuity of Ockeghem's mastery of polyrhythms and how the application of different metres on the same tone material affect the metrical feel of the individual voices. Note that in the Renaissance the metre does not affect the smaller note values below the minim. *Semiminim* and *fusa* continue to have the same length for all four voices.

The hierarchical pattern of subdivisions in Table 3.3 shows some interesting properties that demonstrate a general method for the construction of such patterns. Due to space constraints we can only present the *Discantus*, the highest voice in Ockeghem's composition, but the principle can be easily adapted for the other voices. Starting with the highest metrical level (I) in Table 3.3, which has a period of 6 minims, we see that the largest denominator is equal to the number of mensurations required to finish the meta-cycle together with all the other voices. All other denominators on level I are divisors of this number. The next level (II), with a period of two minims, shows us how many subdivisions are contained in the meta-cycle on that particular

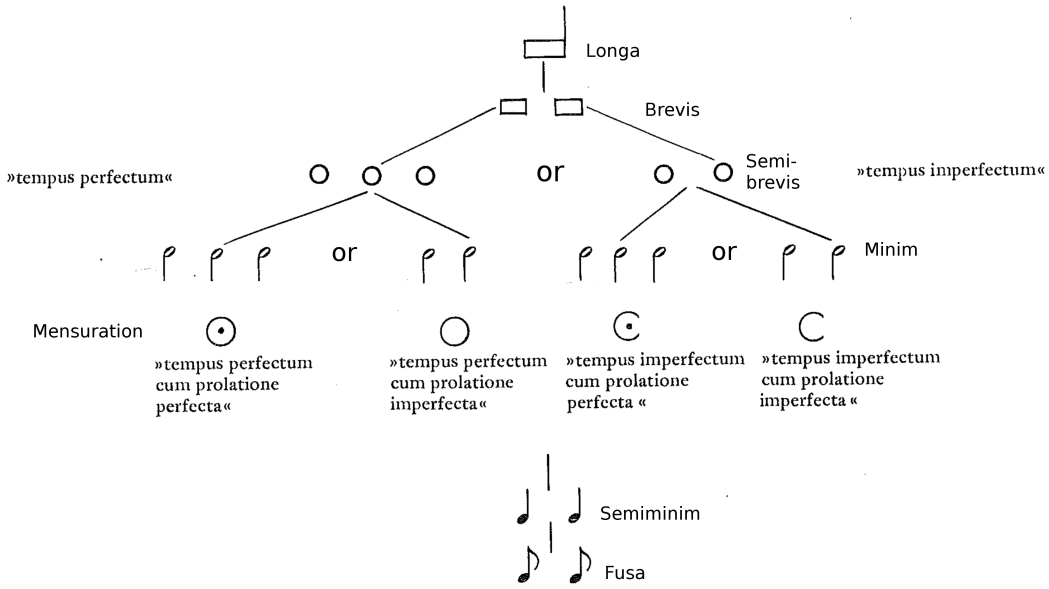


Figure 3-13: System of subdivision used in Renaissance music, beginning with the *ars nova* in the 14th century (de la Motte, 1981). Reproduced with permission.

level. This is again indicated by the largest denominator of the level and all other denominators are its divisors, only *excluding* those already contained on higher levels. This principle repeats itself on all metrical levels. For each of the denominators b in F_{36}, their numerators also follow a simple principle: they traverse the ordered set of numbers coprime to b. In our example, for $b = 18$, its numerators a, in the order of their appearance in F_{36}, are $\{1, 5, 7, 11, 13, 17\}$. This is rooted in the general properties of a Farey Sequence.

In general, rhythmic concepts that depend on a regular pulse deal with various degrees of regularity and irregularity that are applied to durations, tone colours (pitch and dynamics) and accentuation (dynamics and articulation). We would further like to distinguish between repetitive regular patterns and repetitive irregular patterns. The former being obtained by selecting a single element out of F_n and repeating it unchanged in cycles, or by filtering F_n in such a way that only those fractions remain that represent equal inter-onset times. If this process is varied after each cycle then it turns into a pattern of irregular duration contrast. Contrasts that can also be achieved using other musical parameters: accentuation, tone-colour, chord density and dynamics. Repetitive irregular patterns on the other hand are created through a series of *different* elements taken from F_n, for example by using a random selection process. Repetition of this pattern may occur after each cycle. If the series changes after each cycle then this process turns into an irregular duration contrast. Again, one may substitute duration with any other musical parameter.

Stravinsky Igor Stravinsky's music to the ballet *The Rite of Spring*, which was first performed in 1913, contains a famous polyrhythmic passage called the *Procession of the Sage* (Stravinsky,

1967, pp.62). Figure 3-14 shows six layers of pulsations that are associated with different melodic and harmonic material. These are being played by different instrumental groups. The proportions of these pulsations are 12:8:6:4:3 in relation to the length of one bar in 6/4

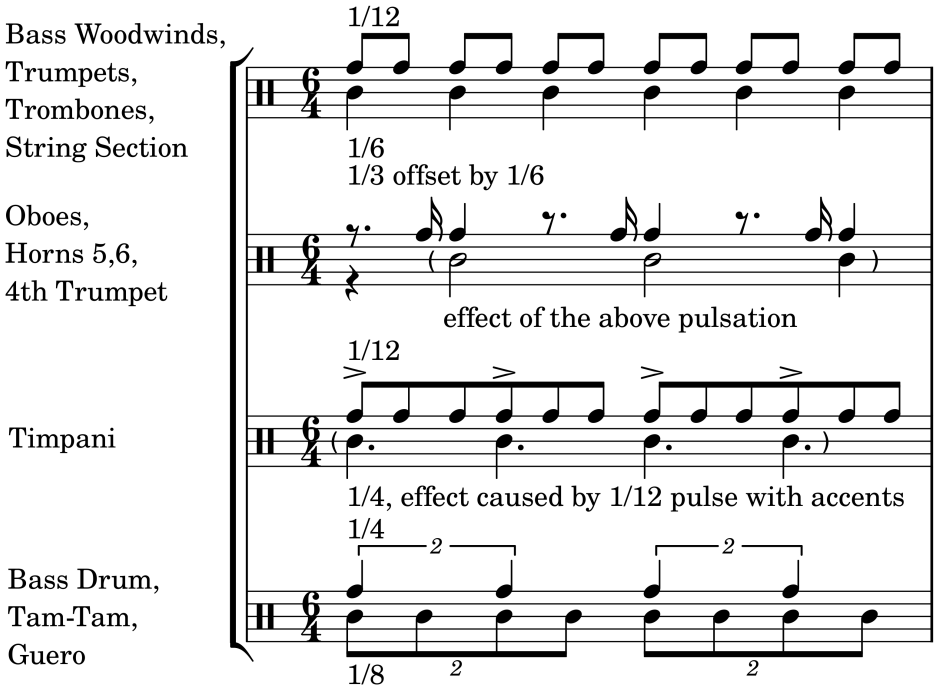


Figure 3-14: Polyrhythmic passage by Stravinsky (1967, pp.63) at rehearsal number 70. Pulsation lengths are given in integer fractions relative to the length of one bar.

metre. When looking at the layering of the melodic material one detects even longer periods for superimposed melodic patterns, which are repeated over and over again in a so-called *ostinato*, see figures 3-15 - 3-19.

The climax of the *Procession of the Sage* is a 48 crotchets long passage at rehearsal number 70 (Stravinsky, 1967, pp.63), where all pulsations start to sound together. They complete one full cycle of six 6/4 bars and then end abruptly with a general silence. Figure 3-20 shows the lengths of the different periods of pulsations and cyclic patterns that Stravinsky employs over six complete bars in a 6/4 metre.

From this analysis we can derive how a Farey Sequence can model the rhythmic structure of the polyrhythms in the *Procession of the Sage*. The compound rhythm in figure 3-14 can be represented by F_{24}, which is filtered in such a way that only the subdivisions by 24, 12, 8, 6, 4 and 3 prevail. The subdivision by 24 ensures the inclusion of the semiquavers of the Oboe pattern. However, due to the participation of the subdivision by 8 in the Guero pattern, only two semiquavers are not covered by any of the other pulse onsets. They have the onset values 11/24 and 19/24. The filtering of F_{24} can thus be refined to allow only the subdivisions

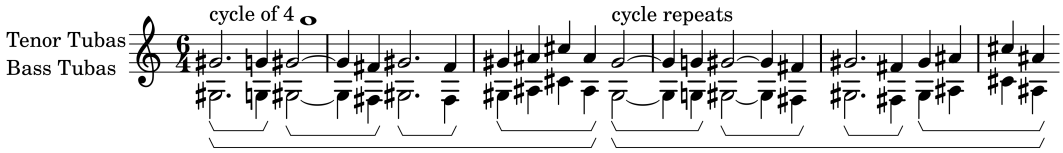


Figure 3-15: The tubas are playing the longest pattern.

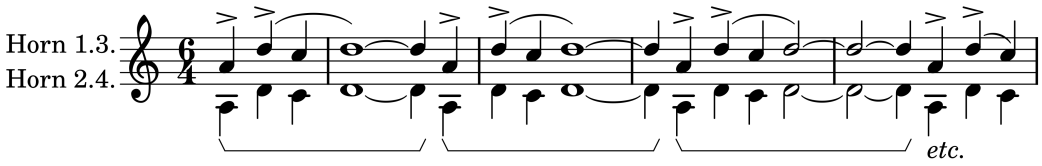


Figure 3-16: The horns have a pattern with an underlying metre of 3+5 crotchets.

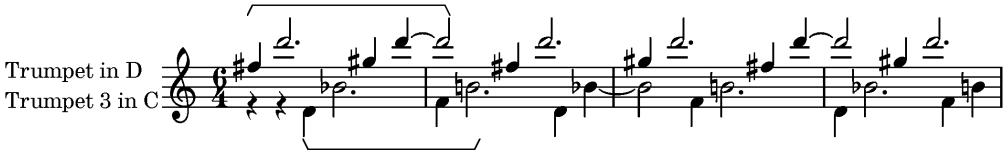


Figure 3-17: The trumpets and trombones play a canon-like pattern that extends for two semibreve. The rhythmic alternation of the two voices is called a *hocket*.

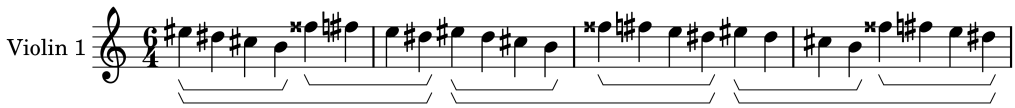


Figure 3-18: The violins and flutes play a descending scale pattern, which repeats after 8 crotchets. The pattern is subdivided into 2 groups of 4 crotchets each.

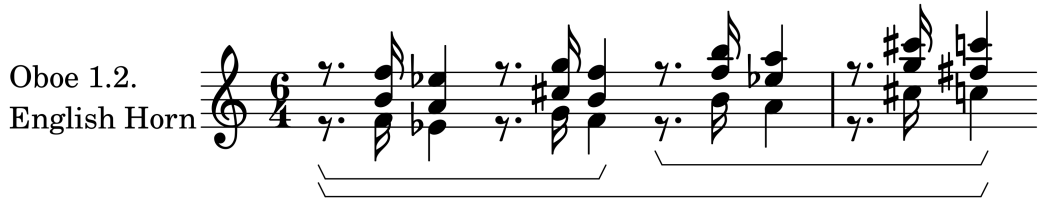


Figure 3-19: Oboes and brass players have an ascending chord pattern that features an off-beat structure due to the rests. The length equals 8 crotchets.

by 12, 8, 6, 4 and 3 with subsequent addition of 11/24 and 19/24; note that both numerators are prime. All denominators in the filtered Farey Sequence are 3-smooth. From the series of 3-smooth numbers up to 12, which is $\{1, 2, 3, 4, 6, 8, 9, 12\}$, only the subdivision by 9 is missing, therefore the only power of three that is allowed to participate is 3^1. In musical terms, the subdivision by 9 would have been a pulsation in crotchet triplets which would have coincided with the normal crotchets at onset times $\{0/1, 1/3, 2/3\}$. Perhaps this would have created an unwanted metrical emphasis at those ternary points of the metre 6/4. It shows that for the subdivisions involved in musical rhythms, not only the prime number factors are important, but also their exponents are equally important. In essence, this polyrhythm turns out to have the structure of a super-hemiola, because we find the superimposed proportions of $2/3, 4/6, 8/12$. Note that all voices participate in one of these hemiolas and that 1/2 is the only point, apart from the downbeat, where *all* pulses from each voice must coincide.

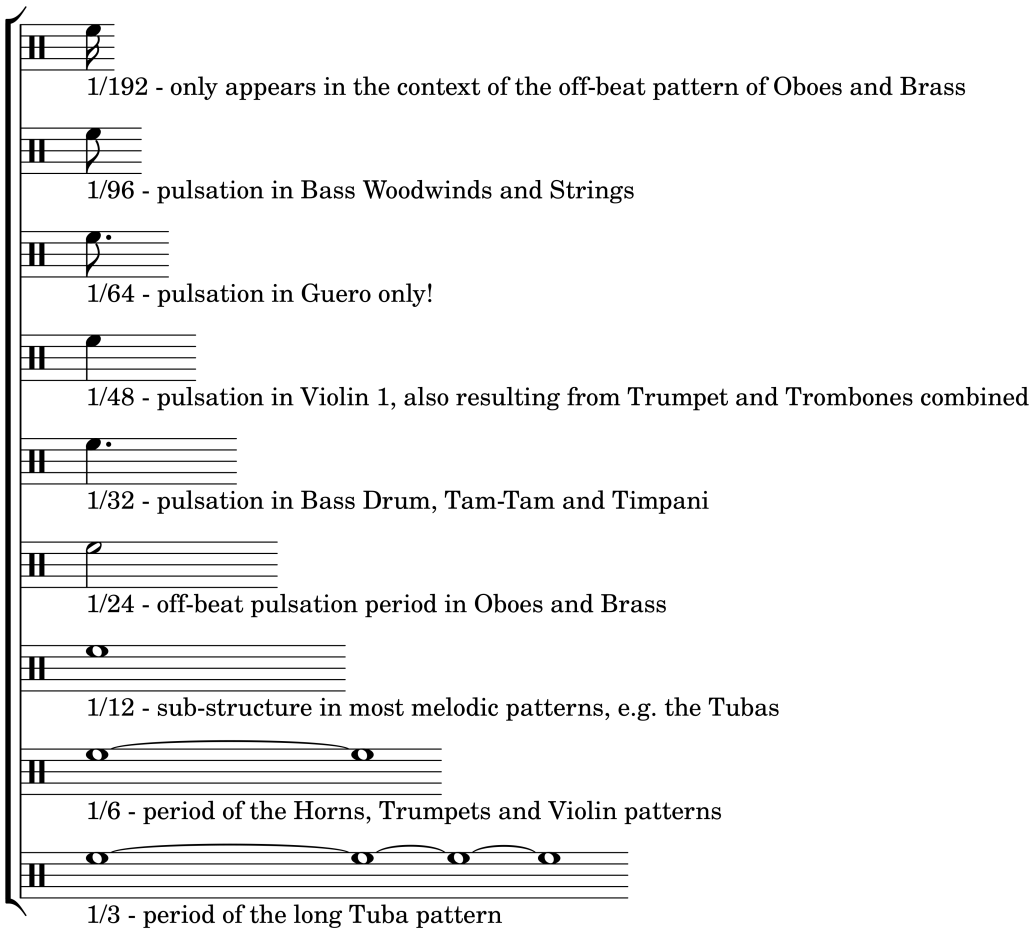


Figure 3-20: Overview of proportions used in the *Procession of the Sage*. Ratios are related to the length of 6 bars in 6/4, i.e., 48 crotchets.

In Stravinsky's second recording of this section, the period for a crotchet is approx. 386 milliseconds, or roughly 155 BPM (Stravinsky, 1960). Based on this tempo we can analyse how the perceptual timescales map to the voices involved in the polyrhythm. The following tempo-metrical grid in figure 3-21 shows the analysis. The extraordinary result is that Stravinsky uses

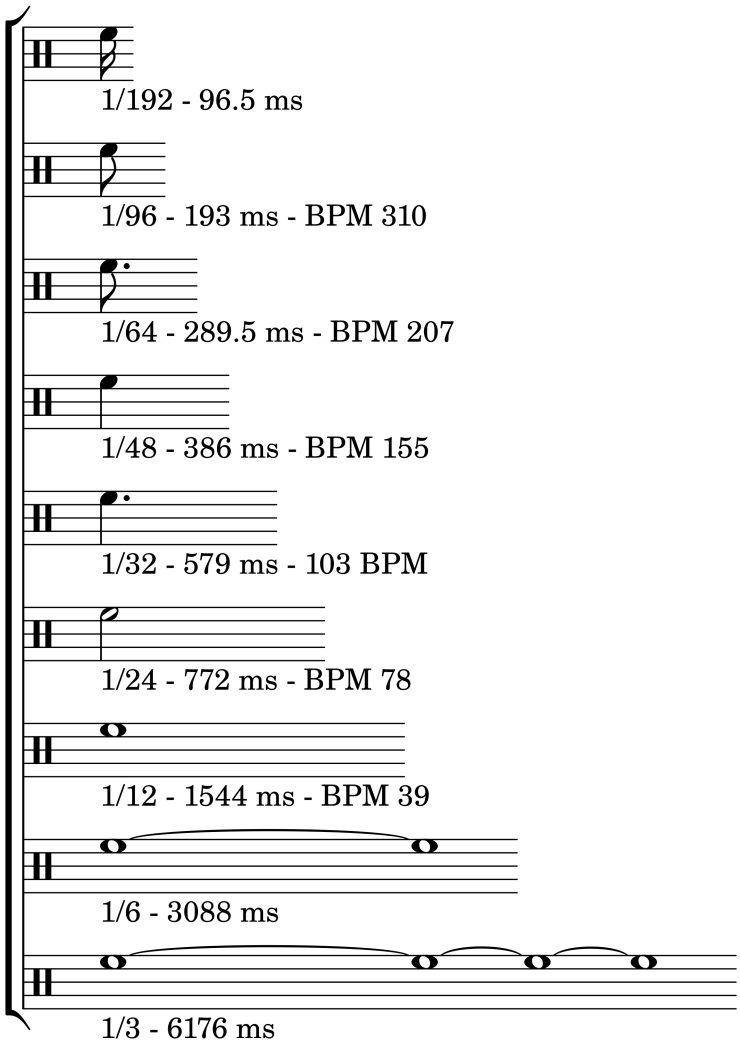


Figure 3-21: The tempo-metrical grid of Stravinsky's own performance of the *Rite of Spring*, section *Procession of the Sage*. See also figure 3-20.

the entire bandwidth of musical timing structures ranging from roughly 100 ms for the smallest subdivision up to slightly above 6 seconds for the largest ternary bar structures (see section 2.4 and especially figure 2-9 on page 45). At the centre we find one period that is exactly in the range of the maximum pulse salience (597 ms), which is the pulsation carried out by the combination of Bass Drum, Tam-Tam and Timpani. This is a pulsation that goes against the

grain of the melodic information in the violins, trumpets, horns and tubas, which is relying on the crotchet subdivision (386 ms). Both streams together participate in one of the hemiola structures (6:4 in relation to one bar of 6/4 metre) that we have discovered earlier.

African Drumming Sima Arom's study (Arom, 1991) has been very influential on contemporary Western composers because of his successful recording and transcription processes that form the basis for his further analysis. We would like to translate some of the principles described in his book into the realm of Computer Music for creative purposes but also in order to demonstrate the general use of our concept. We are referring here to Arom's classification of rhythmic processes (p.203) that form part of the African drumming traditions he investigates but which also occur in some of our Western traditions, see figure 3-22. We would like to present the following 'recipes' in order to create those principles by means of Computer Music.

Type a *Identical durations with regular accentuation*: 1) choose from F_n a series of Ratios with equal distance or delta time d. 2) choose an integer multiple of d for accents and create tables encoding steps 1) and 2).

Type b *Identical durations with irregular accentuation*: 1) choose from F_n a series of Ratios with equal delta time d. 2) choose an integer multiple i of d for accents. Choose i from a table of small integers. 3) repeat step 2) after i beats, thereby creating an irregular accentuation contrast.

Type c *Identical durations with no accentuation but regular alternation of tone colour*: 1) choose from F_n a series of Ratios with equal delta time d and decide on the series' length. 2) according to the F_n frame in 1), map an equally sized list of tone colours, where repetitions may also occur.

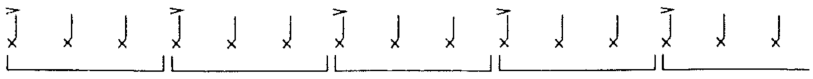
Type d *Identical durations with no accentuation but irregular alternation of tone colour*: 1) choose from F_n a series of Ratios with equal delta time d and decide on the series' *maximum* length in beats. 2) according to the F_n frame in 1), map lists of tone colours not larger than the series in 1), repetitions of tones are allowed.

Type e *Differing durations with regular accentuation*: 1) choose a F_n and filter to obtain desired rhythms, for example via smooth numbers. 2) Place accents within this series at equal delta times.

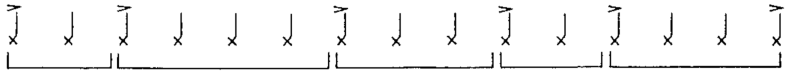
Type f *Differing durations with irregular accentuation*: 1) choose a F_n and filter to obtain desired rhythms, for example via smooth numbers. 2) Place accents within this series at non-equal delta times.

Type g *Differing durations with no accentuation but regular alternation of tone colour*: 1) choose a F_n and filter in such a way that a large subdivision, e.g., 1:2 always occurs. 2) map tone colours in such a way that a change in tone colours only occurs at the boundaries of those large subdivisions.

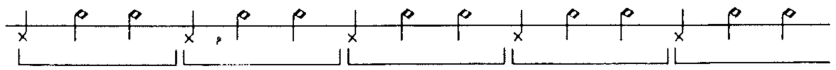
Type a Identical durations with regular accentuation



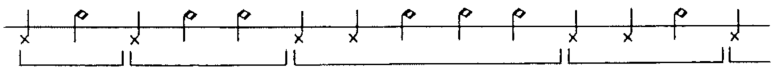
Type b Identical durations with irregular accentuation



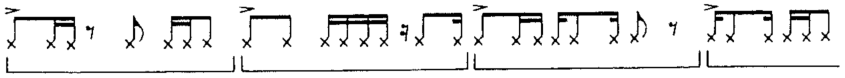
Type c Identical durations with no accentuation but *regular* alternation of tone colour



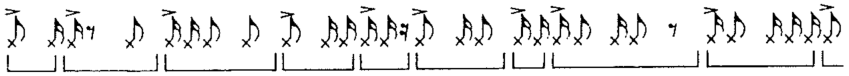
Type d Identical durations with no accentuation but *irregular* alternation of tone colour



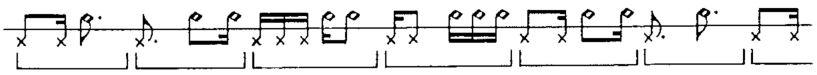
Type e Differing durations with regular accentuation



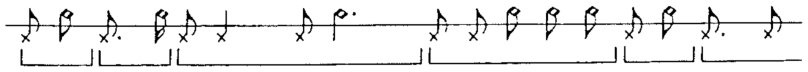
Type f Differing durations with irregular accentuation



Type g Differing durations with no accentuation but *regular* alternation of tone colour



Type h Differing durations with no accentuation but *irregular* alternation of tone colour



Type i Differing durations *with neither accentuation nor change of tone colour*

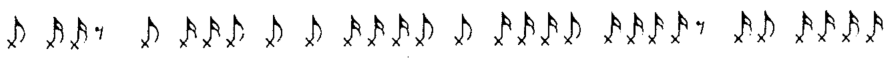


Figure 3-22: The table of the prevalent rhythmic procedures in African music (Arom, 1991). Reproduced with permission.

Type h *Differing durations with no accentuation but irregular alternation of tone colour*: 1) choose a F_n and filter to obtain desired rhythms, for example via smooth numbers. 2) map tone colours in such a way that a change in tone colours occurs at non-regular delta-times.

Type i *Differing durations with neither accentuation nor change of tone colour*: 1) choose a F_n and filter to obtain desired rhythms, for example via smooth numbers or probabilistic filters.

Mix carry out types a) to i) in succession and in polyphony, also apply and mix those methods with regard to other musical parameters.

3.3.3 Rhythm Transformations

Although there are endless possibilities to change a rhythmic motive we are able to classify the principles of variation in the following self-explanatory list, see figure 3-23. We also map the onsets of rhythmic events to elements of the Farey Sequence representing the entire length of the motive. Note that only the operations *Ellipse*, *Augmentation* and *Diminuition* preserve the initially given proportions between note durations. Even more complex shifts of proportions will arise from the simultaneous combination of selected transformations leading to a rich palette of variations. Simple yet musically powerful rhythmic motives, which can serve as the basic cells for these kinds of variations, can be taken from the repertoire of Greek verse rhythms.

3.3.4 Greek Verse Rhythms

Many scholarly works on musical rhythm and music analysis, from Franco of Cologne to Igor Markevitch (1983), make references to Greek Verse rhythms. They are regarded as rhythmic nuclei of larger musical structures. The hierarchy of Western musical structures at least for tonal music develops from the smallest element, the motive, to phrases, sections, movements and entire cycles of movements. The smaller units of form especially are often being explained with reference to Greek Verse rhythms. Because they establish quantitative relations between long and short durations in speech, they were easily picked up for rhythmic analysis within the Western tradition. Figure 3-24 shows a few very common examples of Greek verse rhythms.

We can show how to generalise the creation of instances of verse rhythms by using Farey sequences. In order to build a Iamb or a Trochee we generalise the fact that the second note is shifted away from the centre (0.5) of the time frame (0...1), either towards the first note (0) or towards the end of the time frame (1). The former will produce a Iamb, the latter creates a Trochee. Similarly, in order to create an Anapest we would first create a Iamb, then subdivide the emerging frame between the first and second note of the Iamb. A Dactyl is created by subdividing a Trochee's frame between its second note and the end of is time frame. Note, that the subdivision again may produce a Iamb or a Trochee instead of a clean 1:2 ratio. An Amphibrach is created by the merger of a Iamb and a Trochee. Many more rhythms can be created in a similar way, all of them can be represented by F_n. We can show that every rhythm produced this way can be quantised and rendered using F_n in arbitrary precision. For example,

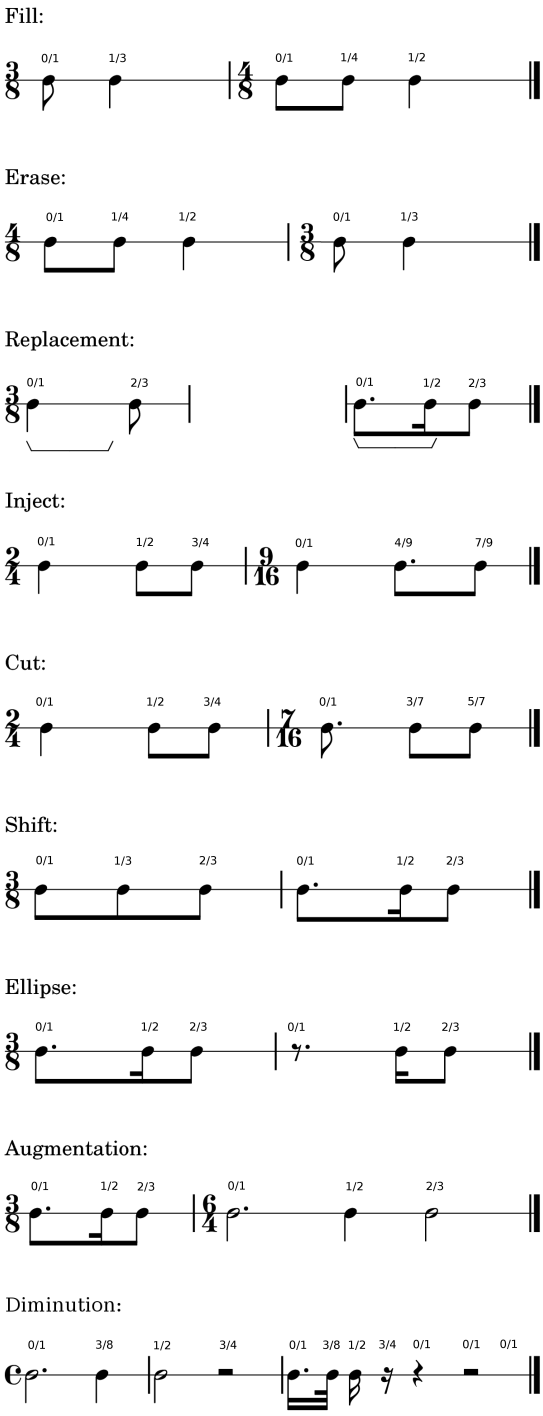


Figure 3-23: List of possible operations in order to create rhythmic variations of small rhythmic motives. The original rhythm is displayed on the left-hand side, its transformation is shown on the right-hand side. Onsets are marked by elements of a Farey Sequence whose length equals the total duration of the rhythmic cell.

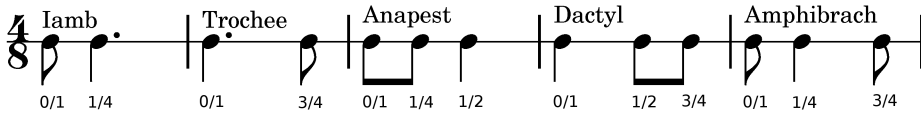


Figure 3-24: Greek verse rhythms in CPN and with onsets mapped to Farey Sequence F_4.

if one wants to work with rhythms that are based on short-long patterns that match precisely a grid of n beats, then, given a number of n beats per bar, the number of possible distinct patterns is 2^{n-1}.

The following figure 3-25 represents our systematic approach for the generation of all Greek verse rhythms and their variants. Sources for the names and for some of the rhythmic structures are taken from Dupré (1925) and Markevitch (1983). Note the new aspect of our representation. For example, the Iamb rhythm can manifest itself under various durational proportions.

If rhythm is defined as the alternation of long and short note durations, then these kinds of rhythms can also be found within Farey sequences. If we let a short duration be a and the long duration be b, then the following sets of short-long-rhythms can be built:

ab	aab	aaab	aaaab	etc.
abb	abbb	abbbb	abbbbb	etc.
aabb	aabbb	aabbbb	aabbbbb	etc.
aaabbb	aaabbbb	aaabbbbb	aaabbbbbb	etc.
etc.	etc.	etc.	etc.	etc.

Each set includes also permutations of the above patterns.

When we look at all 4-beat rhythms, where 'short' is defined as a quaver and 'long' is a crotchet, then the 4-beat rhythms can be built entirely from the sets *ab*, *aab*, *aaaa* and *bb*, as illustrated in figure 3-25.

Similar tables for all 3-beat, 5-beat and 6-beat rhythms, which are formed according to long-short patterns, lead to the conclusion that they can be seen in a more generalised view as filtered Farey sequences. The ratios that are drawn from a sequence F_n are all multiples of the ratio $\frac{1}{n}$ in $[0, \ldots, 1[$ excluding 1. For example, within F_4 we find the list for the 4-beat patterns: $\{\frac{0}{1}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, see equation 3.6.

$$F_4 = \{\frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}\} \tag{3.6}$$

It appears that, based on a list of ratios drawn from F_n, all rhythmic patterns, which belong to the n-beat group are generated by a combination of those ratios, where $\frac{0}{1}$ is always picked as the first element of the beat-pattern. One can identify several aspects of this combinatorial problem. First, the number of possible combinations C_n that form all patterns of n beats can be written in the form

$$C_n = \sum_{k=0}^{n-1} \binom{n-1}{k},$$

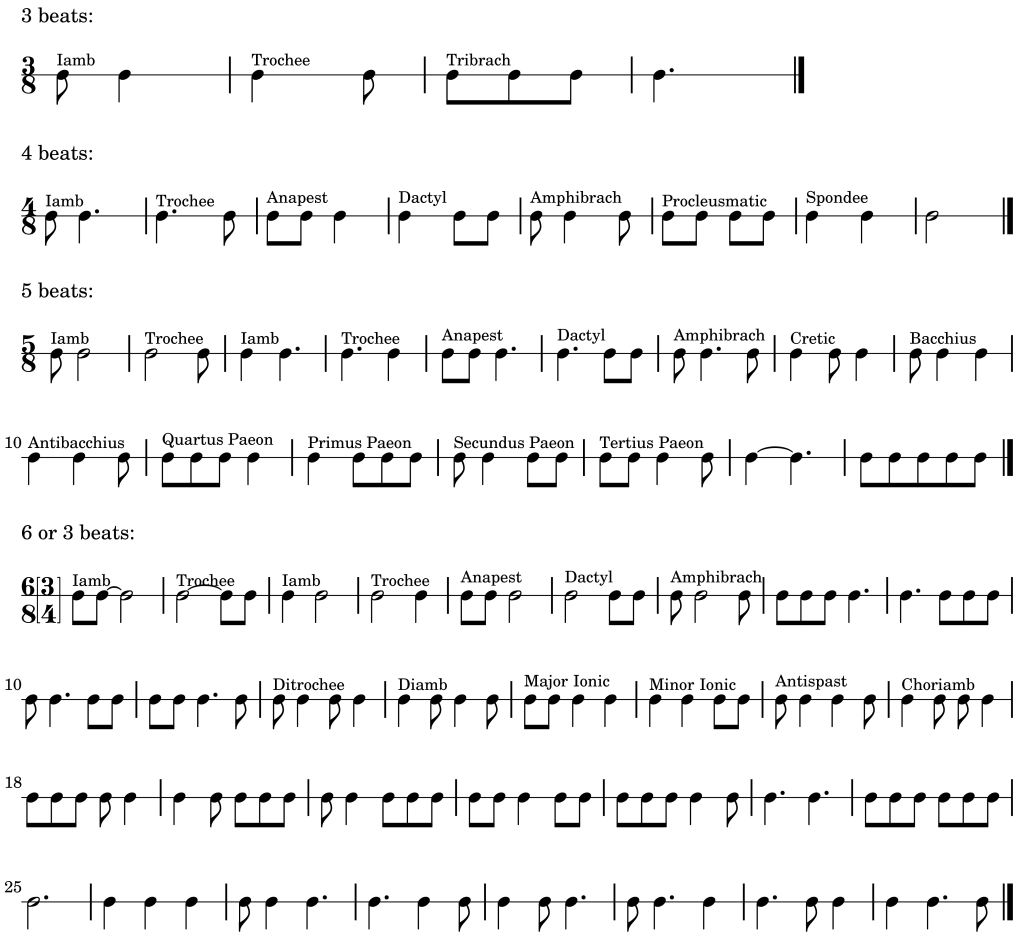


Figure 3-25: Systematic chart of all Greek verse rhythms. The chart can continue infinitely as the number of beats per bar grows with higher $n \in \mathbb{N}$.

which is equal to the sum of the $(n-1)^{th}$ line of the Pascal triangle. Looking at the series of binominal coefficients, they clearly indicate how many patterns belong to a group of patterns that contains an equal number of elements. For example, the 5-beats table contains $\sum_{k=0}^{4}\binom{4}{k} = 16$ different patterns that belong to 5 different groups of patterns. We have therefore $\binom{4}{0} = 1$ pattern of set c, $\binom{4}{1} = 4$ patterns of set ab, $\binom{4}{2} = 6$ patterns of the sets aab and its complement abb, $\binom{4}{3} = 4$ patterns of the set $aaab$ and $\binom{4}{4} = 1$ pattern of the set $aaaaa$. If one looks at the indices of the ratios within the complete list starting with 0 as the first element it becomes clear that in order to construct all patterns algorithmically one can leave apart the 0 index for the moment because the ratio $\frac{0}{1}$ is always the first element in a pattern. Yet it is possible to concentrate on the remaining indices taking them as the set $S_n = \{1, \ldots, n\}$ and performing all combinations in lexicographical order using all possible subsets of S_n. Finally, the omitted ratio $\frac{0}{1}$ is prepended to each one of those permutations plus that ratio appears solo once.

3.3.5 Filters Based on Sequences of Natural Numbers

A musical metre can be described in terms of the distances between peaks of attendance on the beat level expressed as a relatively small integer multiplied by some constant value for the period of the of smallest pulsation involved, i.e., what London (2004) calls the N-level. For example, the sequence $\{3, 3, 2, 2, 2\}$ describes a metre with non-isochronous beats on the backdrop of a constant pulsation with the period 1/12 and $N = 12$. Musical rhythms that emerge from an underlying metrical grid show the same property. For example, Olivier Messiaen's non-retrogradable rhythms can also be described as a sequence of small integers, for example $\{2, 2, 3, 5, 3, 2, 2\}$, see figure 3-2, p.76. In fact, any sequence of natural integers, which was algorithmically generated, can be easily translated into a filtered Farey Sequence and such a filtered F_n can then serve as the model for those rhythms.

The method is as follows: From the sequence of natural integers we form an accumulative set, where for each element of it we take the sum of the previous elements in the original sequence starting with 0, for example: $\{3, 3, 2, 2, 2\}$ becomes $\{0, 3, 6, 8, 10, 12\}$. Its last integer, 12, indicates the order of F_n, so that the filtered Farey Sequence becomes:

$$F_{12} \supset \{\frac{0}{12}, \frac{3}{12}, \frac{6}{12}, \frac{8}{12}, \frac{10}{12}, \frac{12}{12}\},$$

and after reducing the terms:

$$F_{12} \supset F_6 \supset \{\frac{0}{1}, \frac{1}{4}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, \frac{1}{1}\},$$

This method can be applied to sequences of integers of arbitrary length. Its outcome is a filtered Farey Sequence that one can always translate into CPN, see section 3.2.4.

Partitioning

The previously mentioned partitioning of numbers can serve as a generating method for integer sequences that are interesting for the creation of rhythms. Here is an example taken from musical practice. Drummers use stick patterns in order to break-down rhythmic sequences and to distribute their tone-colours. Naturally the stick-patterns follow simple partitions of an integer into partition segments of 1 and 2 strokes, for example the para-diddle: RLRRLRLL is a partition of 8 into the integers: $\{1, 1, 2, 1, 1, 2\}$, which models the number of strokes per stick alternating between both hands. Each stroke has the same length although the tempo can change on a larger scale. Accents can be placed anywhere within the pattern. Another example would be a 7 like: $\{2, 2, 2, 1\}$, or RRLLRRL. One realises easily that the evenness of the number of integers in a partition decides whether the same hand starts the next stroke after the completed partition or indeed the other hand. An even number of segments means the same hand that stroke the first unit of the partition will also strike the first unit after it. If the number of segments is uneven then the hands will change, for example in a 5: $\{2, 2, 1\}$, leading to RRLLR, it is clear that the next stroke is going to be L. But of course a player can decide against it thus either changing to a different permutation of the partition or switching to a partition of an altogether higher integer. The para-diddle pattern $\{1, 1, 2, 1, 1, 2\}$ is a permutation of the lexicographical $\{2, 2, 1, 1, 1, 1\}$, or RRLLRLRL.

Because a partition of n is a sequence of integers, one is able to convert it into a filtered Farey Sequence as a subset of F_n. The unfiltered F_n can therefore model all musical rhythms generated by the partition of the integer n. By choosing a particular partition of n one chooses also a particular subset of F_n, i.e., the Farey Sequence can be filtered by any partition of the integer n. Moreover, for musical applications one can choose from all permutations of a particular partition of n, i.e., a change of the order inside the partition, which is seen as a sequence of integers, will also change the resulting musical rhythm.

3.3.6 Filters Based on the Prime Number Composition of an Integer

We have mentioned previously that elements of a Farey Sequence can be filtered out on the basis of the prime number composition of their denominator. A filter that would allow only k-smooth numbers in the denominator can model the normalised onset times of notes written in CPN, where the highest metrical subdivision used in the score equals k. However, as we have seen in the analysis of the *Procession of the Sage*, it is important to take the prime numbers *and* their exponents into account when developing filters for Farey Sequences, which in turn model specific rhythmic and metric structures.

Clarence Barlow developed a weighting function for the analysis of the prime-number composition of a natural integer. He invented two algorithms that measure the suitability of certain integer ratios to be used in the design of musical tuning systems: The function of "indigestibility" of a natural number and a function for the "harmonicity" of an integer ratio. These functions play a major role in Barlow's algorithmic composition program *Autobusk*, prominently used for works like *Çoğliuautobüsişletmesi* and others (Barlow, 1984). Their purpose is

n	$\xi(n)$
1	0
2	1
3	2.66667
4	2
5	6.4
6	3.66667
7	10.2857
8	3
9	5.33333
10	7.4
11	18.1818
12	4.66667
13	22.1538
14	11.2857
15	9.06667
16	4
17	30.1176

Table 3.5: Table of the indigestibility of the first natural numbers.

to facilitate the creation of tuning systems and pitch and chord selection algorithms used for audio synthesis or MIDI-based sound output where individual note frequencies can be tuned with cent value precision, e.g., an equal-tempered semitone is exactly 100 cents wide. The difficulty in developing tuning systems is to find a sequence of small integer ratios that are suitable for the subdivision of a standard interval, for example subdividing the octave, frequency ratio 2/1. One solution could be for example a series of intervals according to the just tuning system $\{1/1, 9/8, 5/4, 4/3, 3/2, 5/3, 15/8, 2/1\}$ (Barlow, 1984; Moore, 1990). We are using both of Barlow's functions but apply them to the 'tuning' of onset intervals that lead to the least complex notation of a recorded performance.

Barlow's Indigestibility Function To arrive at measuring the "harmonicity" of an integer ratio, Barlow developed his measure of "indigestibility" of a positive integer, $\xi(N)$, based on the number of prime factors involved and by taking into account both the size of each prime number and the size of its exponent, see equation 3.7.

$$\xi(N) = 2\sum_{r=1}^{\infty}\left\{\frac{n_r(p_r-1)^2}{p_r}\right\} \tag{3.7}$$

with $N = \prod_{r=1}^{\infty} p_r^{n_r}$, p_r is the rth prime number and n_r is its exponent in the prime number composition of N. Table 3.5 shows the development of $\xi(n)$ for the first 17 integers.

When sorted after increasing values of $\xi(N)$, the first 17 natural integers have the following order:

$$\{1, 2, 4, 3, 8, 6, 16, 12, 9, 5, 10, 15, 7, 14, 17\}$$

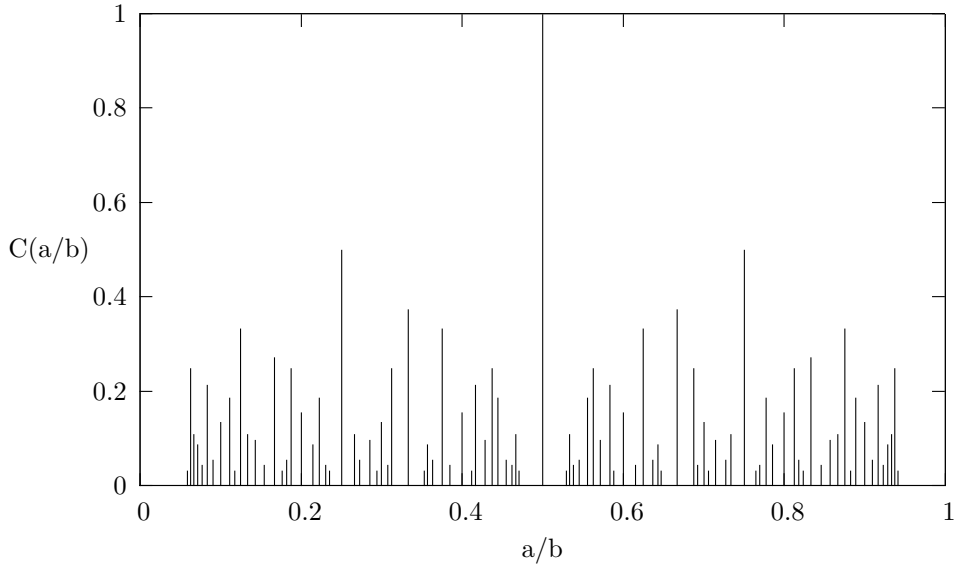


Figure 3-26: Correlation of $a/b \in F_{17}$ and $C(a/b)$ in the interval $]0 \ldots 1[$

We use this measure to determine a weight C of an integer fraction in the context of inter-onset-intervals (IOIs). The indigestibility of the denominator determines the metrical relevance of the fraction representing the onset of a musical event and the same can be done for the IOIs. For any given ratio a/b with $a, b \in \mathbb{N}$ we calculate:

$$C(a/b) = \frac{1}{\xi(b)} \tag{3.8}$$

with $\xi(b)$ defined in equation 3.7. Figure 3-26 shows the correlation between ratios $a/b \in F_{17}$ and their weighted metrical relevance $C(a/b)$ according to equation 3.8.

Barlow's Harmonicity Function Barlow uses his formula of indigestibility to measure the harmonicity $H(p/q)$ of a pitch interval p/q according to:

$$H(p/q) = \frac{sgn(\xi(q) - \xi(p))}{\xi(p) + \xi(q)} \tag{3.9}$$

Barlow uses the signum function in order to take the root position of the interval into account. For example, in terms of musical harmony, the fundamental of the perfect fourth, $H(3/4) = -0.214286$ is the upper note, whereas in the perfect fifth, $H(2/3) = 0.272727$, it is the lower note. The higher $|H(p/q)|$, the higher is the subjective stability of the interval (Hajdu, 1993). Therefore a tritone tuned as 45/64 has only a harmonicity of 0.056391. Because our research focuses on the transcription of IOIs the polarity of those ratios are not interesting for us,

therefore we simplified the equation 3.9 so that it always returns a positive number:

$$H(p/q) = \frac{1}{\xi(p) + \xi(q)} \tag{3.10}$$

with $H(p/q) = 1$, if $p == q$. The harmonicity measure has proved to be useful in order to evaluate a set of quantised IOIs in preparation for their transcription into CPN. The problem of inferring a beat structure underlying a set of durations, based only on knowing the durations, could be solved with the help of the harmonicity function in equation 3.10. We will point out the details of it in section 6.2.1 on page 141.

Euler's gradus suavitatis Barlow's ideas are related to Euler's *gradus suavitatis* measure that had been also applied to musical pitch relationships. In an attempt to measure the "grace" and aesthetic pleasure of certain pitch constellations Euler proposed this measure in a letter to Johann Bernoulli, dated May 25 1731 (Cerrai et al., 2002, p.286):

> Let many different notes be taken whose *numeri pulsuum* [numbers of beats], that occupy the same number of time, stand to each other as the whole numbers a, b, c, d etc., by which the same notes are usually expressed. Let A be the *minimus communis dividuus* [lowest common multiple]. I call this number the *exponentem* of the same notes, because it is on this base that we recognise grace that is produced when the same notes are played either at the same time or in succession. I thus devised a *gradus suavitatis*, the first of which includes the most perfect chord , that is to say, when all notes are relatively equal. The following ones include the less perfect, according to their order. From the *exponente* it is possible to recognise the *gradus suavitatis* in the following manner: I break it down into its *factores simplicissimos* [most simple factors], I add these together, and then I subtract from their sum $n-1$ (n stands for the *numerum factorum* [number of factors]); the number that is thus obtained represents the *gradum*. For example, let the notes be 1, 2, 3, 4, 5, 6, then the *exponens* of these notes will be 12. The *factores simplicissimi* of this are the following three: 2.2.3, the sum of which, less 2, gives 5. It is clear from this that the harmony of these notes will be pleasant to the 5^{th} degree. In this way, it is possible to find the *exponentem* of a whole piece of music, if all the notes are expressed as whole numbers and the *minimum commune dividuum* is taken.

The algorithm described by Euler can be written as:

$$G(N) = \sum_{r=1}^{\infty}(n_r p_r) - \sum_{r=1}^{\infty} n_r + 1 \tag{3.11}$$

with $N \in \mathbb{N}$, p_r is the rth prime number and n_r is its exponent in the prime number composition of N. If the algorithm wants to measure the *gradus suavitatis* of a list of natural integers, then one has to build the lowest common multiple of the list first. We tested the *graudus suavitatis* on the Farey Sequence F_{17} where each of the ratio's denominator is measured. Figure 3-27

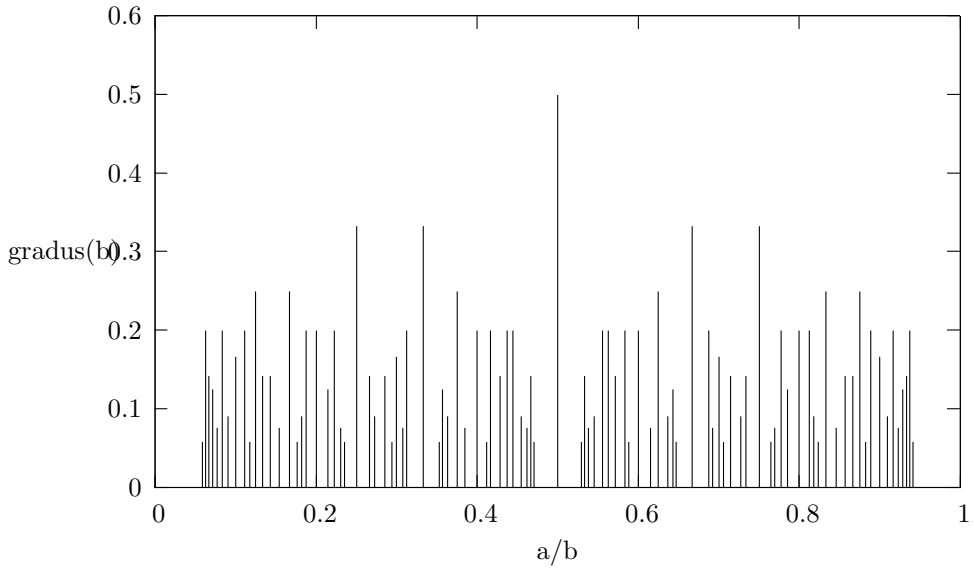


Figure 3-27: Correlation of $a/b \in F_{17}$ and *gradus suavitatis* $1/G(b)$

shows the position of each ratio $a/b \in [0 \ldots 1]$ on the x-axis while the inverse *gradus s.* $1/G(b)$ is plotted on the y-axis. If you compare this outcome with the weighting measurement of the ratios after equation 3.8 you see that the latter produces a more differentiated outcome. For example a number of values $1/G(b)$ are equal which could lead to potential ambiguities when used as a criterion for quantisation.

3.3.7 Metrical Filters

Due to its nature discussed earlier, a filtered F_n can represent an entire bar structure from isochronous or non-isochronous groups of beats down to all possible levels of subdivision and beyond towards meta-bar structures. Therefore, a filtered Farey Sequence can be used to create the dot notation commonly used within music analysis to denote the various layers of metrical structures (Lerdahl and Jackendoff, 1996; London, 2004).

Calculation of the Metrical Tempo Grid London (2004) found interesting links between psychological thresholds of time perception in music listening and the prevalence of certain metrical subdivisions and bar structures in Western music, see section 2.4. He uses a systematic tree structure to visualise beat subdivisions in multiples of 2s and 3s and metres that also employ groupings of beats into units of 2s and 3s. Those factors of subdivision and metric grouping are therefore taken from the sequence of 3-smooth numbers. The periods of those are then mapped on the basis of a main beat period. Due to the psychological thresholds, this tree exhibits areas of possible subdivisions and metres given a specific tempo of the beat. Bars that become too long and subdivisions, which are too short are then cut from the tree. The remaining tree maps

out a metrical tempo grid, see table 2.1, p.57. The decision to use only multiples of 2s and 3s is based on the assumption that the majority of Western music to date is covered in this way. Subdivisions involving prime numbers above 5, and metres employing higher primes than 3 were indeed rarely used in Western music until approximately 100 years ago, when the 20th century avantgarde music started to test and push these boundaries. In order to reflect this development, one can expand London's proposed tree structure easily by using higher primes for metric subdivisions and beat groupings.

The following steps describe how to use the Farey Sequence for the calculation of the metrical tempo grid. The outcome of our method is identical with the metrical tempo grid described by London (2004, pp.38), see also table 2.1, p.57.

1. Calculate F_{27}, because the highest integer for metrical subdivision is 3^3 for the majority of Western music.
2. Filter F_{27} so that only those integer ratios remain that have 3-smooth numbers as denominators.
3. From the filtered F_{27} keep only the reciprocals.
4. Duplicate every reciprocal in this set and take the inverse from the duplicated fractions.
5. Multiply every fraction in this set by a fraction that represents the duration of the beat, for example $\frac{1}{4}$.

Table 3.6 shows, as an example, the resulting set of integer ratios representing time signatures when they are greater than $\frac{1}{4}$, otherwise they represent subdivisions of the beat. $\frac{1}{4}$ is set as the beat with a period of 650 milliseconds. The perceptual thresholds now apply, and durations below 6 seconds and above 100 ms are emphasised.

Graphical Derivation of Metrical Hierarchy It is possible to derive metrical hierarchies from a filtered Farey Sequence F_n. Subdivisions of a musical duration generate an isochronous IOI, i.e., the ratio between the duration and the number of subdivisions n. Metrical hierarchies based on the subdivision by n emerge graphically from the tree of F_n, because integer fractions with a denominator smaller than n are rooted on a higher level of subdivision, see figures 3-29 and 3-30. Here, each of the higher levels would then represent a different beat cycle, and the level of subdivision by n corresponds to London's N-cycle (London, 2004). For example, F_6 can represent an N-cycle of 6 subdivisions in a bar. The resulting 6 IOIs of duration $\frac{1}{6}$ can either be interpreted as a metre with 2 beats and 3 subdivisions (time signature $\frac{6}{8}$), or they can be interpreted as a metre with 3 beats and 2 subdivisions (time signature $\frac{3}{4}$). This interpretation depends on how the musical context is structured and perceived. The 2 beat cycle is marked by the fractions $\{\frac{0}{1}, \frac{1}{2}\}$ and the 3 beat cycle is marked by the fractions $\{\frac{0}{1}, \frac{1}{3}, \frac{2}{3}\}$. The Farey Sequence is therefore able to represent the hierarchic structure of isochronous metres (London's I-metre). These are metrical structures that have isochronous beat periods. Metres with varying beat periods are called non-isochronous metre (NI-metre). Note that only divisiors of n can form

duration or metre	value (float)	period [ms]	BPM	note name
1/108	0.00925926	24		
1/96	0.0104167	27		
1/72	0.0138889	36		
1/64	0.015625	41		
1/48	0.0208333	54		
1/36	0.0277778	72		
1/32	0.03125	81		demisemiquaver
1/24	**0.0416667**	**108**	**554**	semiquaver triplet
1/16	**0.0625**	**163**	**369**	semiquaver
1/12	**0.0833333**	**217**	**277**	quaver triplet
1/8	**0.125**	**325**	**185**	quaver
1/4 [beat]	**0.25**	**650**	**92**	crotchet
2/4	**0.5**	**1300**	**46**	minim
3/4	**0.75**	**1950**	**31**	dotted minim
4/4	**1**	**2600**	**23**	semibreve
6/4	**1.5**	**3900**	**15**	dotted semibreve
8/4	**2**	**5200**	**12**	breve
9/4	**2.25**	**5850**	**10**	3 dotted minims
12/4	3	7800		dotted breve
16/4	4	10400		
18/4	4.5	11700		
24/4	6	15600		
27/4	6.75	17550		

Table 3.6: A metrical tempo grid after London (2004, p.44) with the central beat set at 650 milliseconds or 92 BPM. This table was generated by our method using Farey Sequence F_{27}. Emphasised are all durations and metres within the range of perceptual timing limits, i.e., between 0.1 and 6 seconds.

a beat-cycle and only this relationship between N-cycle and beat-cycle generates an I-metre. Therefore, prime-numbered N-cycles will always result in an NI-metre but can never become an I-metre. In order to model an NI-metre on the basis of F_n, with a metrical subdivision by n as the N-cycle, one would use partitions of n to mark the starting points of beats within the N-cycle. The implementation of the above metrical hierarchies uses a subdivision filter in order to extract all members of F_n that are related to the same level of subdivision, for example the subdivision by 6, shown in figure 3-30, extracts from all F_n, with $n \geq 6$, the fractions $\{\frac{0}{1}, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, \frac{1}{1}\}$. The common IOI between these onsets is of course $\frac{1}{6}$.

There are well-formedness constraints for the construction of NI metres which affect the application of partitions to an N-cycle. According to WFC 5 stated earlier, there needs to be at least one time point of the N-cycle left between consecutive beats. This excludes all partitions of N that use '1' as an element. Figure 3-28 shows an example of a 12-cycle NI-metre with *5* beats as found in London (2004, pp.127). Each of the nodes represents a member of F_{12} that takes part in the metrical subdivision of $\frac{1}{12}$. Possible NI-metres emerge from the application of partitions of 12 while taking into account WFC 5 and also WFC 6 with regard to maximal evenness of the metre. A list of possible NI-metres of the 12-cycle using those partitions is

shown in table 3.7. Interestingly, the partition 3-3-2-2-2 that is the basis of Bernstein's *America* rhythm, is the only kind of partition of 12 with no '1', with only prime numbers used and with 3 as the highest prime number involved. Variants of this metre are formed by permutations, e.g., 2-2-3-3-2. Higher powers of two and three on the beat level do not emerge because they automatically become subdivided into twos and threes in line with WFC 6.

As we have pointed out earlier in section 3.3.5, any partition of an integer, which is a sequence of integers, can be converted directly into a filtered Farey Sequence.

3-3-2-2-2
2-3-3-2-2
2-2-3-3-2
2-2-2-3-3
3-2-2-2-3
3-2-3-2-2
2-3-2-3-2
2-2-3-2-3
3-2-2-3-2
2-3-2-2-3

Table 3.7: Partitions of 12, which are usable for a 12-cycle NI-metre; all of them contain 5 beats.

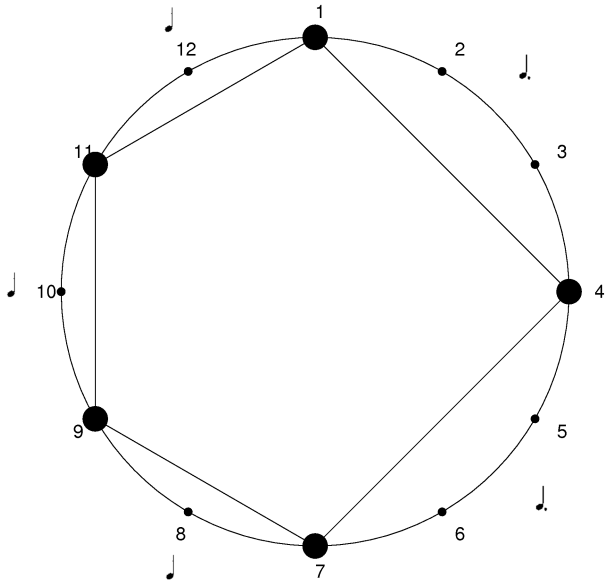


Figure 3-28: A 12-cycle NI-metre with 5 beats models Bernstein's "America" rhythm.

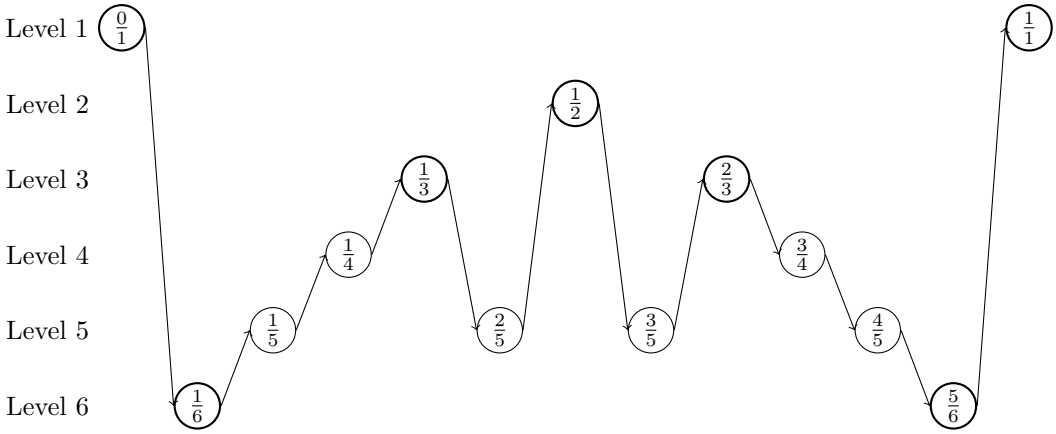


Figure 3-29: Farey Sequence F_6. Fractions a/b are placed on their respective bth level.

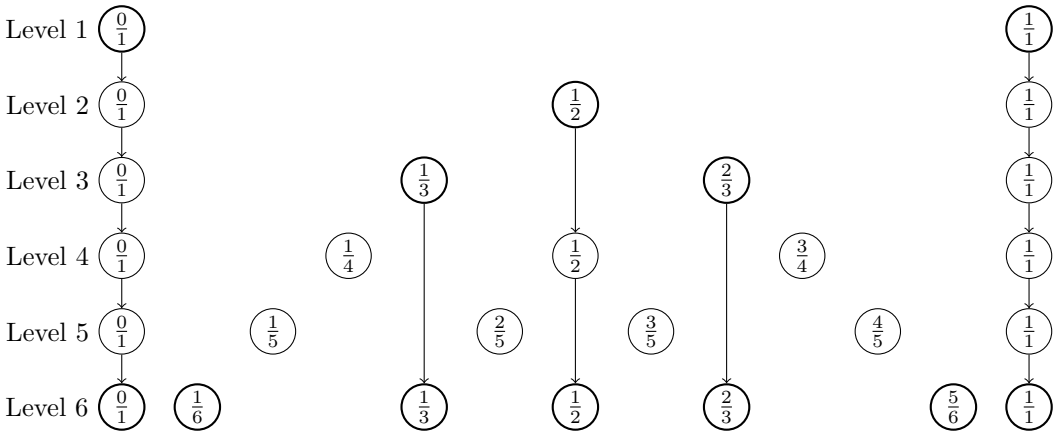


Figure 3-30: Graphic derivation of metres 3/4 and 6/8 from Farey Sequence F_6. The 2nd level is the beat level for the 6/8 metre. The 3rd level is the beat level for the 3/4 metre. Both metres have the potential to form hemiolas by placing accents on the beats of the neighbouring levels 2, 3 or even 4. The 6th level is the isochronous level of pulsation, or N-cycle (London, 2004).

3.4 Summary

The aim of this thesis is to show that musical rhythm and metre can be modelled by using filtered Farey Sequences. The Farey Sequence is particularly useful because it can be scaled into different ranges of human temporal perception and varying temporal dimensions of musical form. A filtered Farey Sequence can model beat subdivisions and patterns of musical metre. With it we can also analyse large musical structures, as for example Igor Stravinsky's *Procession of the Sage*, see section 3.3. Or we can chain larger sequences of Farey Sequences together, which form then a single filtered Farey Sequence of higher order, see section 3.2.

The Farey tree, which is a branch of the Stern-Brocot tree, is also used for the analysis of

mode-locking behaviour in weakly coupled oscillators. Large and Kolen (1995) studied their use for beat-tracking but have not arrived yet at a complete working system.

We are making use of the Farey Sequence in a much broader sense and cover many aspects of musical timing, rhythm and metre. We have shown that we can arrive at a filtered Farey Sequence by converting arbitrary sections of CPN. This process is reversable, see section 3.2.4.

The connection between perceptual timing thresholds and musical metre can be explained through so-called metrical tempo-grids (London, 2004). We have demonstrated that it is possible to model these metrical tempo-grids with the use of an algorithm that is based on the Farey Sequence F_{27} filtered by 3-smooth numbers. The algorithm can be easily extended in order to include higher prime factors for beat subdivisions and musical metres, see section 3.3.7 and table 3.6. We have also shown that a Farey Sequence can graphically describe metrical hierarchy. Thus a Farey Sequence can generate the dot notation commonly used for metrical analysis.

We are able to show that rhythms and metres, as well as large sections of a musical piece in CPN can be analysed and explained by filtered Farey Sequences. We gave a wide range of examples from different cultures. Polyrhythms from the simple 3:2 proportion of the *hemiola* to much more complex examples can all be found in Farey Sequences. The analysis of Greek verse rhythms, proportional canons from the Renaissance, African Drumming, Messiaen's non-retrogradable rhythms, any rhythm encoded as an integer sequence, highly ornamented Baroque music, in all of these cases we can operate with Farey Sequences and we are thus able to generalise the method of filtered Farey Sequences for music analysis. In addition, the analysis of a polyrhythmic passage from Stravinsky's *Rite of Spring* could also show a link between Stravinsky's own recording of the piece and the perceptual timing limits that have been reviewed by London (2004), and which form the basis of his theory of musical metre. The cyclic patterns in necklace notation that have been used as well for metrical analysis can be easily reproduced by using a filtered Farey Sequence.

We have also worked within a team at the University of Bath in order to incorporate a rhythm model based upon the Farey Sequence into a system for algorithmic composition, ANTON, which allows its users to compose pieces of music according to a user-defined set of rules and in accordance with the programming paradigm of *Answer Set Programming*, see Boenn et al. (2008) for an overview.

Chapter 4

Experimental Framework

4.1 Introduction

In this chapter we would like to set out how we are going to test the main hypothesis of this thesis: Are filtered Farey Sequences capable of modelling musical rhythm and metre? As we have seen in the previous chapter, it is possible to model rhythm and metre with filtered Farey Sequences on the basis of scores written in CPN and of metres or cyclical rhythmic patterns represented in necklace notation or as sets of natural integers. We have also seen that the metrical tempo grid proposed by London (2004) can be modelled algorithmically by using a filtered Farey Sequence. In the rest of the thesis we are interested in the question whether the Farey Sequence serves equally well in the quantisation and transcription of real-world musical performances. We are specifically interested in quantising performed onset times of compound rhythms. For example, a piano score involves two hands playing in polyphony. It is their combined action, the compound rhythm of their performance, which hits our ear. This perceptual content, which consists of perceived onsets, is sent to subsequent neural processes that are involved in recognising learned patterns, such as musical metre, and also rhythmic patterns that are composed over an underlying musical pulsation or metre. We have seen in section 2.3 that the transients of sounds carry most of the crucial timing and rhythm information. Therefore we are specifically interested in the analysis and transcription of onsets recorded from musically performed compound rhythms.

We will introduce now our test material and propose a distance measurement, which helps to determine the quality and success of the quantisation and transcription process.

4.2 Test Material

We have chosen two recordings of the *Aria* of the *Goldberg Variations* by J.S. Bach played by Glenn Gould in 1955 and in 1981. Onsets in both recordings were extracted using the algorithms of the *aubio* library by Brossier (2006). A certain amount of hand editing was necessary to delete wrongly detected onsets or in order to place an undetected onset marker. Both onset

streams were segmented into analysis windows by hand. We have tested the quantisation and transcription algorithm by using windows that were aligned to the length of one bar in 3/4−metre.

The Bach *Aria* is rhythmically interesting because it contains many ornaments, which have only been written in shorthand notation, for example trills, mordants, appogiaturas etc.. This means that these parts of the piece are not written down metrically and that it is left to the performer to interpret these signs and to execute them freely, i.e., not metrically but with expressive timing. It follows that the exact way that a certain ornament is played is not predictable, which is a useful challenge for an automated quantisation and transcription algorithm. We have chosen Gould's performances because they are extremely different from each other, yet it is the same musician performing, which stresses the point that we have made about playing musical ornaments. Gould's playing is extremely interesting for the analysis of expressive timing and its temporal micro-structure (Bazzana, 1997; Oswald, 1997), not the least because of the conscious choices made by Gould to contrast his last recording with the early rendering from 1955. The tempo of the *Aria* alone is at times almost twice as slow in 1981. In addition to the abundance of different ornaments, the *Aria* features syncopated rhythms, room for showing expressive phrasing, all in contrast with a smooth sequence of semiquavers in the last quarter of the piece.

We also tested the quantisation and transcription program with artificially generated sequences of onsets.

4.3 Distance Measurements

As part of the quantisation process, the quantised sets of durations are analysed in terms of their distance from the performed durations. We also included an analysis of the distance of the quantised set from the original section of score. We have evaluated the outcomes of our quantiser by using the following Euclidean distance measure, see Toussaint (2004) for a review.

4.3.1 Euclidean Distance

Given two n-dimensional vectors of durations (IOIs),

$$X = \{x_1, x_2, x_3, ..., x_n\}$$

and

$$Y = \{y_1, y_2, y_3, ..., y_n\},$$

the Euclidean distance $d_E(X, Y)$ is measured by:

$$d_E(X, Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4.1}$$

Clearly, when the distance between a quantised set of durations and the original set of durations within the score is equal or very close to zero, we should see a perfect or very close match between them, which is the aim of our automated quantisation system. We can also extract viable information from the Euclidean distance between the quantised set and the performed set of durations. We would not expect a zero distance, of course, because then we would not have quantised anything. However, staying as close as possible to the performed durations might be beneficial and a useful feature of the program.

4.4 A Measure for Contrast in Rhythmic Sequences

We would also like to introduce a measure for contrast in rhythmic sequences described by Patel (2008). The normalised pairwise variability index (nPVI) has been used in studies of speech rhythms of various languages to show how theses rhythms correlate with a language's category of being either stress-timed or syllable-timed. Low et al. (2000) first developed this index to investigate the patterning of syllable durations. The nPVI "measures the degree of contrast between successive durations in an utterance" (Patel, 2008, p.131), and it is calculated by the following equation (p.177):

$$nPVI = 100/(m-1) \times \sum_{k=1}^{m-1} |(d_k - d_{k+1})/((d_k + d_{k+1})/2)| \tag{4.2}$$

with m = number of durations and d_k = the duration of the kth element. For every subsequent pair of durations we measure values between 0 (both durations being equal) and 2 for the maximum durational contrast, i.e., one of the durations approaches zero. The nPVI of a sequence of IOIs with at least two elements is therefore limited between 0 and 200. Given any sequence of durations, higher numbers of nPVI indicate a greater degree of durational contrast within the sequence, whereas lower nPVI values are characteristic for sequences where large contrasts between durations do not occur. Therefore, the nPVI is a good tool for making comparisons between individual duration classes of a performed musical sequence. For instance, it allows one to study the level of timing contrast within a class of performed quavers. The nPVI has also been used in comparative studies of speech and music rhythms where their results suggest that the spoken language of a composer has an influence on the rhythmic structure of her or his music, see Daniele and Patel (2004). We find the nPVI useful in order to have a comparative measure of rhythmic smoothness applicable to the sets of duration classes, i.e., clusters of durations that are similar in length, and the original sequence of durations. It is expected that, by comparison, duration classes show a relatively low nPVI value, and if there is more than one duration class, the value of nPVI of the entire sequence should be significantly higher than the nPVIs of each of the duration classes. In the next chapter, we will show how the duration classes are established from recorded performances.

Chapter 5

Grouping of Onset Data

5.1 Introduction

We dedicate a chapter to the grouping of onset times and inter-onset-intervals because this technique plays a crucial role in our automated quantisation program. If listeners categorise musical rhythms on the basis of simple integer ratios (Desain and Honing, 2003; Papadelis and Papanikolaou, 2004), and if they are also capable of following smooth tempo variations and correlate the rhythmic content successfully, then we assume that there must be a kind of window following mechanism that buffers short sections of inter-onset-times and another process groups similar inter-onset-times into duration classes, whose relation to each other is represented by a quasi-simple integer ratio. When changing the tempo smoothly those duration classes do also shift accordingly in order to keep the sense of a stable rhythmical structure. One can regard this phenomenon as a manifestation of the Gestalt-principles of 'similarity', 'good continuation' and 'common fate' (see section 2.7 for an overview of Gestalt principles). Rhythmic stability is defined as the manifestation of simple ratios mainly between performed event onsets. The stability is maintained in the performance of those ratios although the absolute timing values of the duration classes change during an accelerando or ritardando. Because of this, a duration class can only be meaningful when it can be related to other duration classes in the temporal vicinity with whom specific ratios of durations can be established. This means that although there are variations within each duration class, the mean ratio between different ratio classes should remain more or less constant on a local timescale although their absolute timing values can change significantly. We expect also border-line cases where one inter-onset-interval (IOI) might belong to two different duration classes at a time. We will discuss how our transcription algorithm will handle these ambiguities.

In this chapter, we will describe how IOIs can be grouped into duration classes starting with an ordered set of onset times extracted from a musical performance. This grouping algorithm works on windowed sets of onset data. Apart from splitting onset data into analysis windows by hand, we will also present two different automated windowing mechanisms that produce a set of analysis windows. The classification of IOIs is carried out per individual window. We

have found a metric that allows for the clustering of similar IOIs. We have also developed an algorithm, which arrives at a unique representation of duration classes per analysis window.

5.2 Calculation of Duration Classes

We analyse an ordered set of note onsets S from a musical performance. S can be generated from MIDI note-on messages, see section 2.3.2, or via onset detection from audio recordings, for which we have used the *aubio* library (Brossier, 2006).

$$S = \{s_1, s_2, s_i, ..., s_{|S|}\} \tag{5.1}$$

Our approach to the quantisation of S by filtering out expressive timing, and the subsequent transcription of S into common practice notation (CPN), begins with the classification of IOIs into duration classes according to the following method.

Based on S, an ordered set of inter-onset intervals (IOIs) D is formed:

$$D = \{(s_2 - s_1), (s_3 - s_2), (s_{i+1} - s_i), ..., (s_{|S|} - s_{|S|-1})\} \tag{5.2}$$

The IOIs represent the time interval elapsed between note events, for which we use the term duration. Then, the set D is split up into n analysis windows W, where each window contains a sequence of normalised IOIs:

$$W = \{d_1, d_2, d_i, ..., d_{|W|}\} \tag{5.3}$$

The length of these sets depend on the windowing algorithm that has been used. There are three methods available that apply the appropriate windowing to the set D:

1. Downbeat markers are set in D by hand.
2. An algorithmic segmentation process sets a mark whenever it detects a change from long to short IOIs in D satisfying certain conditions. This mark sets the beginning of a new window W.
3. A sliding window analysis.

Studies on perceptual timing thresholds suggests that, at a very slow tempo, the performance of a bar becomes 5-6 seconds long, see (London, 2004), which is also the time that has been suggested as the limit of psychological present (James, 1950; Pöppel, 1972; Michon, 1978; Fraisse, 1984). Therefore, an analysis window should be at most 6 seconds long and, within such a window, the proposed grouping algorithm is able to cope with performed variations of musical timing. Optimum results for option 2, the automated marking and windowing mechanism, worked on the basis of a minimum window length of 1.3 seconds, i.e., twice the period of the indifference interval of 650 ms (Fraisse, 1963; Wundt, 1911), see also section 2.4.

For each window, a grouping algorithm is performed for each duration: The absolute value of the difference between a single duration and every other duration of the window is calculated one by one: $|d_i - d_j|$ with $i, j = 1, 2, 3, ..., |W|$.

Each absolute value is tested by a threshold $\epsilon \times d_i$, below which we detect a similar duration between d_i and d_j. The threshold varies per single duration, i.e., it varies with the minuend d_i, whereas ϵ is a constant with value 1/6. This constant has been derived from the smallest IOI that is contained in Farey Sequence F_3, namely $\frac{2}{3} - \frac{1}{2}$. We have experimentally found that this value is big enough in order to capture subtle tempo changes but small enough to make sure that those IOIs are not grouped together that would originally take part in two different duration classes, i.e., specifically ternary and binary duration classes. In other words, ϵ is chosen in such a way that triplet and non-triplet forms of durations can be distinguished from each other, whilst making sure that smaller tempo variations within those classes are still tolerated.

If a difference between durations $d_i - d_j$ falls below $\epsilon \times d_i$, the threshold of similarity, then the subtrahend d_j is stored in a set. This set contains in the end all durations d_j close to d_i inside the analysis window 'from the point of view' of a particular duration d_i. We call this set of durations a *duration class* and the procedure is carried out for each duration within the window. The resulting sets are merged if they are similar. Redundant sets or duration classes are deleted and unique duration classes are established for each analysis window. Here is the complete algorithm:

Let the analysis window be a set of normalised durations:

$$W = \{d_1, d_2, d_i, ..., d_{|W|}\}. \tag{5.4}$$

Define sets of durations for $i = 1, 2, ..., |W|$:

$$E_i = \{d_j | (|d_i - d_j|) < (\epsilon \times d_i), j = 1, 2, ..., |W|\} \tag{5.5}$$

with $\epsilon = \frac{1}{6}$. Reduce $\{E_i\}$ so that there are no duplicates. To make the reduction process more efficient, only the integer indices j of the elements d_j are being compared. There are two further rules for the reduction of $\{E_i\}$:

1. If $|E_i| > 2$ and $|E_j| > 2$, then: If $|E_i \cap E_j| \geq 2$, then $E_j \cup E_i$.
2. If $|E_i| \leq 2$ or $|E_j| \leq 2$, then: If $|E_i \cap E_j| \geq 1$, then $E_j \cup E_i$.

This simple mechanism removes sets which carry a relatively low amount of information, whereas the larger sets are having more information to contribute towards the final structure of a duration class. The following table 5.1 shows the groups detected in an example of Glenn Gould's 1955 performance of the very first two bars of the *Goldberg Variations*, see also the score in figure 5-1. The algorithm uses indices of the IOIs of a particular analysis window, because the use of integers facilitates speed and efficiency of the group-list evaluation, i.e., the detection and deletion of redundant lists as well as the detection of the similarity as described

above. Note also that due to those measurements, one single IOI value might be shared between two duration classes. The consequence is that the shared IOI will contribute to the calculation of the centroids (arithmetic means) of both classes involved. It turned out that this allowance improves the results of quantisation algorithm. The final decision to which of the duration classes that shared element will belong to is easy to solve: The class whose mean is closer to the shared IOI will claim it in the end, whereas the other classes will lose this element. It is perfectly possible that a duration class might only have a single element. It can also happen that two IOIs, which, with regard to their notation, would have belonged to the same duration class, are actually being grouped into two different duration classes. This is true, for example, in the second bar of our example shown in Table 5.1: The durations 0.054814 and 0.071365 belong to two different classes, although they have the same duration written in the score. For this case we made sure that the quantisation process, which follows the grouping, has the possibility to assign the same quantised duration, i.e., durations for CPN, to more than one duration class. Table 5.1 shows also an example of how we evaluate the arithmetic means $M(E_i)$ that are taken from every single duration class E_i.

bar	durations	E_i	$M(E_i)$	$M(E_i)$ fraction	prime factors
1	0.328377	0.328377	0.32738	55/168	$168 = 2^3 \times 3 \times 7$
	0.326383	0.326383			
	0.0262919	0.0262919	0.0273798	21/767	$767 = 13 \times 59$
	0.0284678	0.0284678			
	0.230462	0.230462	0.230462	56/243	$243 = 3^5$
	0.0600181	0.0600181	0.0600181	65/1083	$1083 = 3 \times 19^2$
2	0.162128	0.162128	0.164916	47/285	$285 = 3 \times 5 \times 19$
	0.054814	0.175654			
	0.071365	0.156967			
	0.175654	0.054814	0.054814	37/675	$675 = 3^3 \times 5^2$
	0.156967	0.071365	0.071365	80/1121	$1121 = 19 \times 59$
	0.379071	0.379071	0.379071	163/430	$430 = 2 \times 5 \times 43$

Table 5.1: Measurement of duration classes E_i and their mean values $M(E_i)$ from the compound rhythm of the first two bars of Glenn Gould's 1955 performance of the *Goldberg Variations*. The rightmost column shows the prime factors of the denominator of the fraction $M(E_i)$.

An analysis of the integer ratios shows that a further quantisation is needed for musical score transcription. We remember that almost every duration in CPN is effectively based on a ratio whose denominator is a 3-smooth number, see table 3.6, p.111. Because prime factors > 3 in the denominators of $M(E_i)$ are very common, we have to quantise the onsets of the duration classes $\{E_i\}$. Furthermore, we have learned that listeners are entrained to patterns of a specific metrical hierarchy that is based on simple integer ratios. We have found in Barlow's indigestibility function (section 3.3.6, p.105) a tool for producing a weighting measure for onsets that takes into account the relative size of the prime factors of an integer and their exponents. This weight is calculated from the denominator of the normalised onset time represented as an integer ratio within the analysis window W. This weighting measure is useful in order to form an

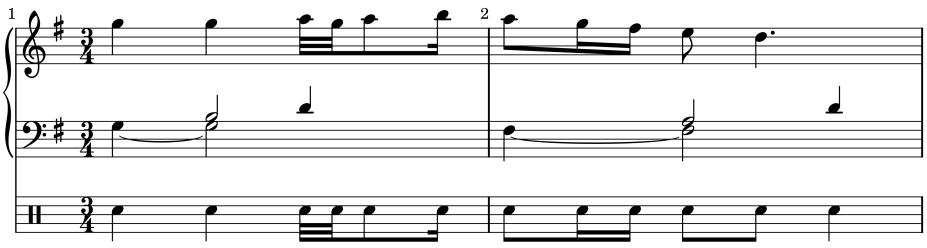


Figure 5-1: The first 2 bars of the *Aria* of the *Goldberg Variations*. The third staff shows the compound rhythm generated from the means of the duration classes, $M(E_i)$ in table 5.1, after further quantisation.

underlying hierarchic metrical grid. Although rhythmic and metrical patterns are dependent on the musical style, metrical subdivisions by 2 or 3 are more likely to happen, especially in Western music (Barlow, 1984; Lerdahl and Jackendoff, 1996; London, 2004). Barlow's indigestibility function yields relatively small weights for integers composed of the prime factors 2 and/or 3 as compared to integers with higher prime factors. By taking the inverse of Barlow's weighting function we can construct a ranking of rational onset positions within CPN. It is the aim of the quantisation algorithm to shift the mean values of the duration classes, $M(E_i)$, towards higher ranking onset positions in CPN.

The following table 5.2 shows how we measured the success of the proposed grouping method. The onset data were extracted from Gould's 1955 recordings of the *Goldberg Variations*. The 32 bars of the *Aria* were analysed separately and the resulting durations classes are compared with Bach's score, but also by taking into account how Gould executes the ornaments notated by Bach. CPN uses a shorthand notation for musical ornaments such as trills. The specific rhythmic execution of the ornament is left to the discretion of the player. We found and overall success of the grouping algorithm with an average error of wrong duration grouping in the range of 10% to 11%.

The next two figures 5-2 and 5-3 show our grouping algorithm working on the first two bars of Bach's *Aria* of the *Goldberg Variations* recorded by Glenn Gould in 1981. We will demonstrate the complete transcription of the *Aria* using also the early Gould recording of 1955 in chapter 6.

5.2.1 Grouping of Noisy Beat Sequences

We would like to find out more about the resilience of the grouping algorithm against random variation of onset timing and how it affects the creation of duration classes. A series of tests consisted of modulating isochronous beat-sequences with noise. We used a sequence of beats with a period equal to the indifference interval of 600 ms (BPM 100). Windows containing 25 beats, or 15 seconds in total length, are modulated by 10% white noise, as for example in figure 5-4. The white-noise modulation is used here to simulate tempo variation in musical performance through unexpected rubato and noise introduced either by the player's manual

bar	error rate (%)	bar length (sec.)	bar	error rate (%)	bar length (sec.)
1	0	2.7575	17	38.1	3.2245
2	16.7	2.8095	18	0	3.6135
3	27.8	2.4205	19	13.3	3.4015
4	27.3	2.9295	20	0	3.524
5	0	2.926	21	11.8	3.457
6	9.1	3.036	22	27.3	3.552
7	7.7	2.9615	23	0	3.901
8	10	3.872	24	10	4.069
9	0	3.106	25	0	3.3905
10	0	3.207	26	7.1	3.401
11	23.8	2.9875	27	0	3.4795
12	12.5	3.329	28	0	3.503
13	11.1	3.5205	29	8.3	3.3875
14	0	3.6545	30	8.3	3.485
15	11.1	3.7325	31	0	3.835
16	11.1	4.2545	32	50	6.232
average error: 10.7 %					

Table 5.2: Measuring the rate of durations that are correctly grouped together based on a comparison with the score and Gould's performance in 1955. We calculate the error in terms of the percentage of durations per bar that have been wrongly assigned to a particular class.

difficulties or originating from his motor apparatus. The same test was repeated with 15%, 16% and 17% noise and was successfully detecting all beats as one duration class. From 18% to 20% white noise content the beats were divided up into two different duration classes with slightly different mean values $M(E_i)$. At 20% noise we found that $M(E_1) \approx 1/25$ and $M(E_2) \approx 1/26$. It appears that the $16.\overline{6}\%$ threshold set for the detection of duration classes, $\epsilon = 1/6$, sets also the limit for the detection of a single class of isochronous beats when they are modulated with such an amount of white noise that it moves some IOIs above the threshold. But, with the mean values of two duration classes close to each other, the quantisation algorithm, which we will explain in the next chapter, is still able to find the isochronous beat sequence, because then the quantiser's search areas around the two mean values do overlap.

There must be a considerable amount of contrast between durations in order to create a sense of rhythm that differs from a continuous stream of pulses that might have smooth *tempo* changes. Such a continuous pulse-stream would be perceived as one that has a varying tempo but not having enough durational contrasts between single events so that it would trigger the notion of different durational qualities. It is remarkable that, on the one hand, there are physically different event durations within a certain time due to changes in tempo but they do not invoke the sense of having a different classifying quality, because the change of duration between adjacent individual events is too little, so that the Gestalt principles of similarity, common fate and good continuation take over and establish the subjective notion of a single duration class although their objective absolute duration is not the same. This psychological effect of tempo change while keeping a constant subjective duration has been also exploited in

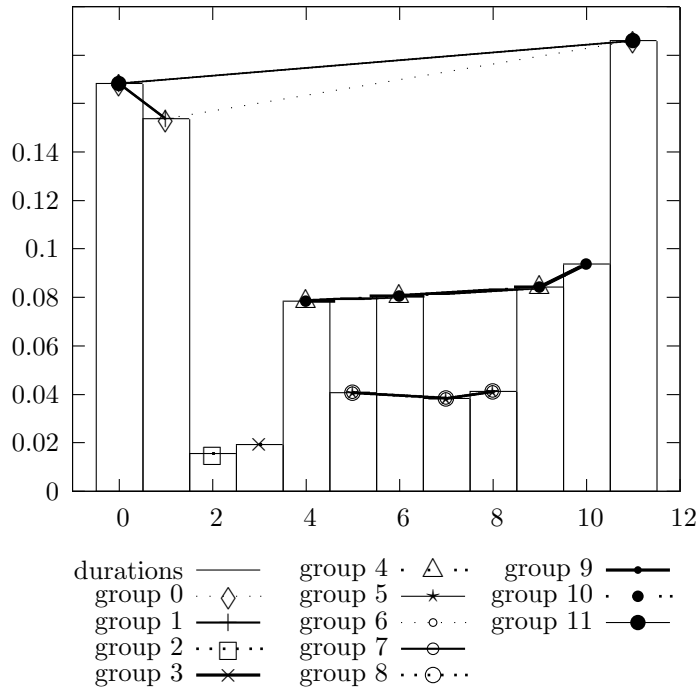


Figure 5-2: This graph presents the normalised durations and the unreduced set of duration classes $\{E_i\}$, as groups 0 - 11, detected in Glenn Gould's 1981 recording of the *Aria* of the *Goldberg Variations*, bars 1-2.

a psychoacoustic paradox very similar to the Shephard tones or the Risset Glissando (Moore, 1990, pp.221). While those two work in the domain of pitch, there is another one associated with Risset that works in the domain of subjective time where sub-audio frequencies of pulse-trains are continuously in- or decreased while being kept at a 1:2 frequency ratio (or within equal distance on a log frequency scale). The amplitudes of the pulse-trains are modulated with a quasi-Gaussian window whose length equals the time-difference between the point reaching the high frequency-threshold and the the point of the low frequency-threshold[1].

5.2.2 Grouping of a Tempo-Modulated Rhythmic Ostinato

Now it would be important to get a notion for the appropriate lengths of analysis windows, so that the grouping into duration classes can perform successfully, also under the circumstance of continuous tempo changes. The test involved a track of 40 bars, 80 seconds long, with an initial tempo of 120 BPM. The track repeats a 2-bars ostinato rhythm, see figure 5-6, that

[1]There have been at least two orchestral pieces in the 20th music that exploited this psychoacoustic paradox: The Variation No. 9 of the *Variations for Orchestra* by Elliott Carter and a large section of the piece *passage / paysage* by Matthias Spahlinger. An early type of composed accelerando has been written by Arthur Honegger in his *Pacific 231*, where note values are continuously shortened to mimic a locomotive's noise of the engine when standing at first and is then set into accelerating motion. Perhaps the composer with greatest ambition and sophistication to experiment with these sorts of rhythmic processes was Conlon Nancarrow.

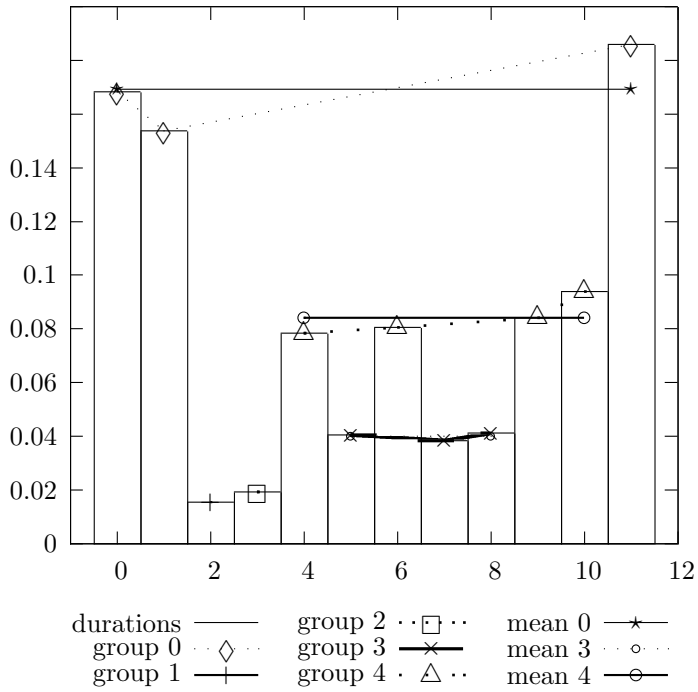


Figure 5-3: This graph represents the final result of the grouping algorithm in Glenn Gould's 1981 recording of the *Aria* of the *Goldberg Variations*, bars 1-2. The nPVI of the original data is 60.7923, group #0 has an nPVI of 13.8958, groups #3 and #4 have nPVIs of 6.46476 and 5.9617 respectively.

is continuously changing in tempo. We tested the 3-2 *son clave* pattern also used by Cemgil (2004) for tempo tracking, see also Kilian (2004) for a discussion. Our track has been modified by a sine wave signal with a period equal to the track length and an *amplitude* = 0.5, i.e the tempo slows down to half of its initial value, then it accelerates until the initial tempo is doubled, and finally it returns back to the inital state, all of it over the course of 40 bars. To track such a big swing of tempo changes is challenging. One cannot expect an accurate tracking of duration classes from a single window that is 80 seconds long. Figure 5-5 shows the IOI data of our test track.

However, we are able to track these tempo changes when we split the 40 bars up into smaller windows, each one may contain up to 4 bars. The reason for this is that the grouping algorithm would break if the analysis windows becomes so long that duration classes would start to overlap. The overlapping of duration classes has to be prevented. The easiest way to avoid overlaps of this sort is to break the performance down into smaller windows, for example each containing an equal number of 4 bars, or 10 IOIs. The following figures 5-7, 5-8, 5-9 and 5-10 present such a solution and show that the grouping algorithm now truly follows the less dramatic tempo changes within the analysis windows.

The challenge of the windowing method lies in finding the right size of the windows to make

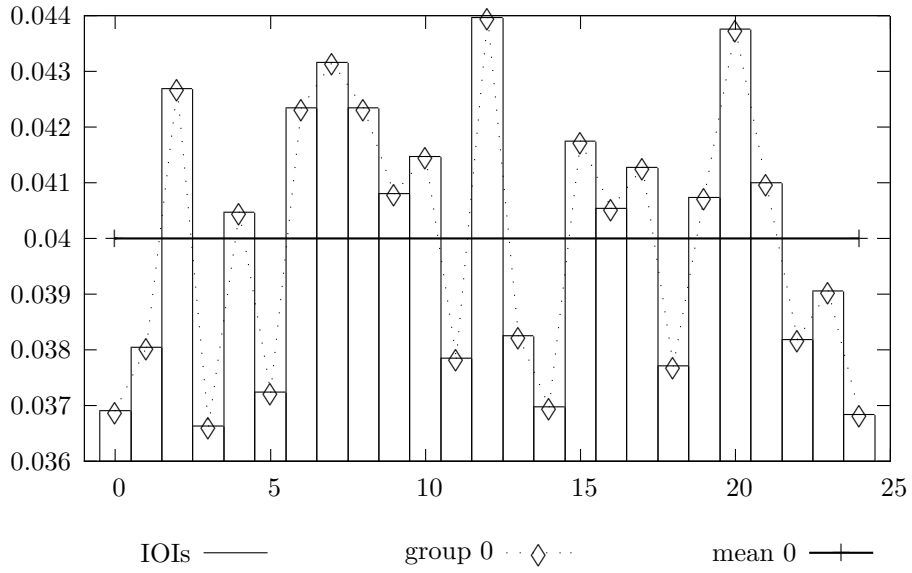


Figure 5-4: 25 isochronous beats of length 600 ms modulated by 10% white noise and normalised. All beats were collected into a single group. The nPVI of both sequences is 7.24648

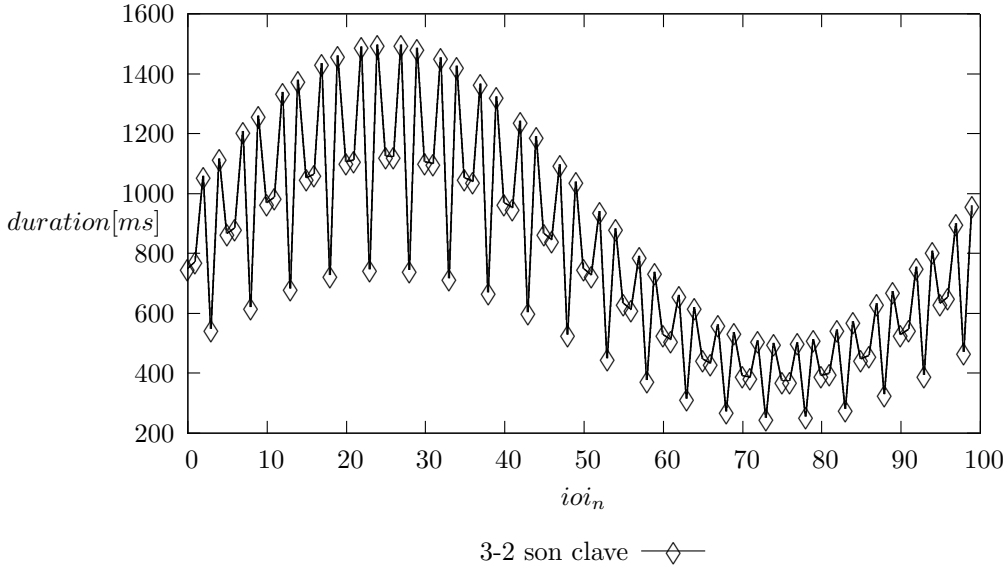


Figure 5-5: 40 bars of a 3-2 *son clave* pattern tempo-modulated by a sine wave. The nPVI of the entire sequence is 38.6457.

sure that the duration classes of the performance are not overlapping prior to the application of the grouping algorithm. One could, for example, take into account the perceptual timing thresholds that have been discussed earlier in section 2.4 and see if these can be usefully applied to the grouping algorithm. In the following sections, we will present two different approaches

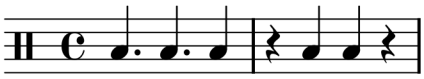


Figure 5-6: The 3-2 son claves pattern.

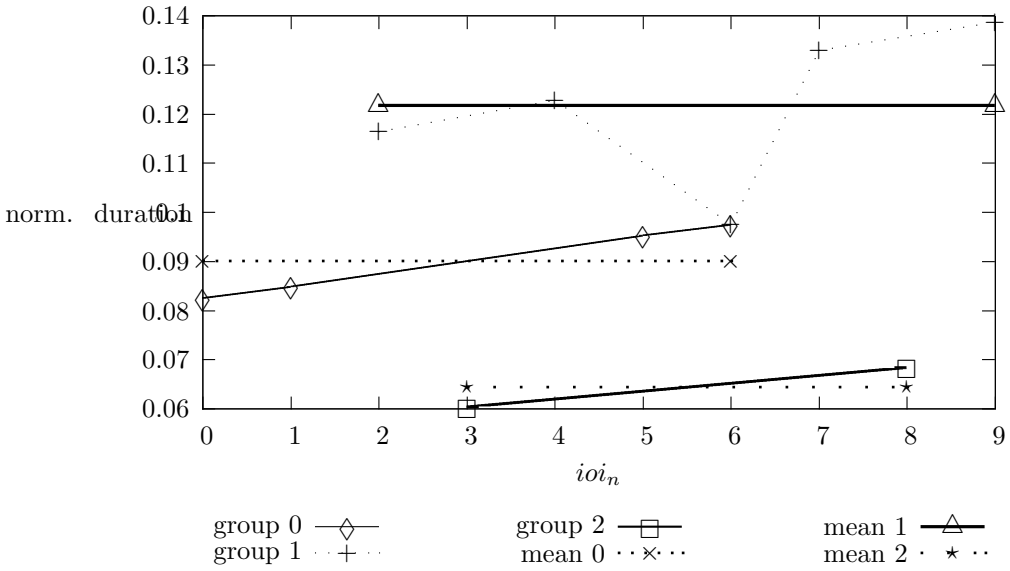


Figure 5-7: First four bars of the 3-2 *son clave* pattern presented earlier. An approximation of the mean values yields: $\{1/12, 1/9, 1/16\}$, which is very close to the original set of IOIs $\{1/12, 1/8, 1/16\}$.

of automatic segmentation of onset data.

5.3 Automatic Onset Segmentation

In order to process greater musical structures automatically like for example entire sections or complete movements of music, we experimented with another approach of windowing streams of onset data. The idea is to measure the ratio of pairs of successive IOIs in order to detect a significant change in rhythm, for example a quaver followed by a semiquaver. This method needs to be general enough to cover a wide range of possible transitions between note durations. Let D be an ordered set of durations:

$$D = \{d_1, d_2, d_i, ..., d_{|D|}\} \tag{5.6}$$

Then for $i = 1, 2, 3, ..., (|D| - 1)$, the value of the ratio $r_i = d_i/d_{i+1}$ will be > 1 if d_{i+1} is shorter than d_i. On the other hand, $r_i < 1$ if $d_{i+1} > d_i$. The algorithm then changes the values of r_i,

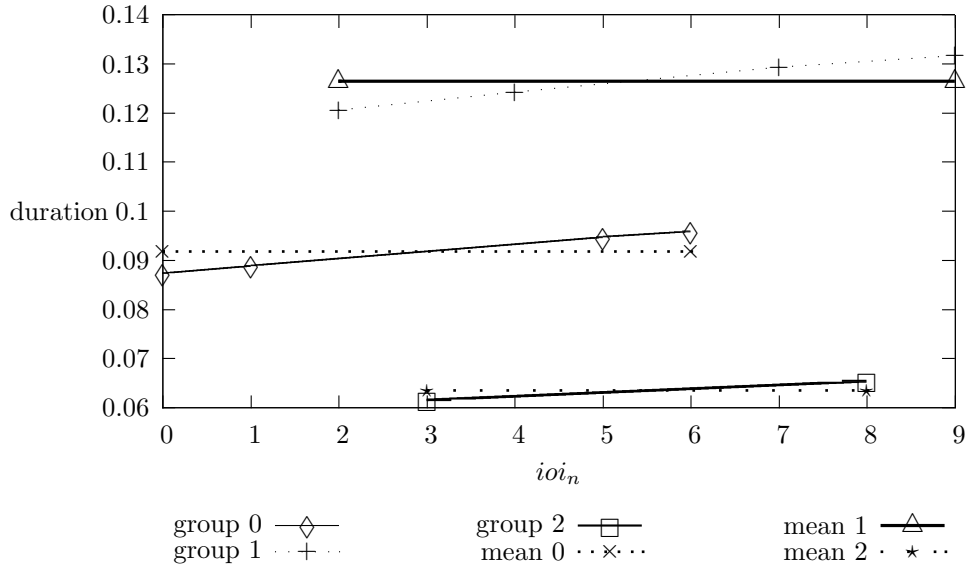


Figure 5-8: Bars 5-8 of the 3-2 *son clave* pattern presented earlier. An approximation of the mean values yields: $\{1/11, 1/8, 1/16\}$, which is very close to the original set of IOIs $\{1/12, 1/8, 1/16\}$.

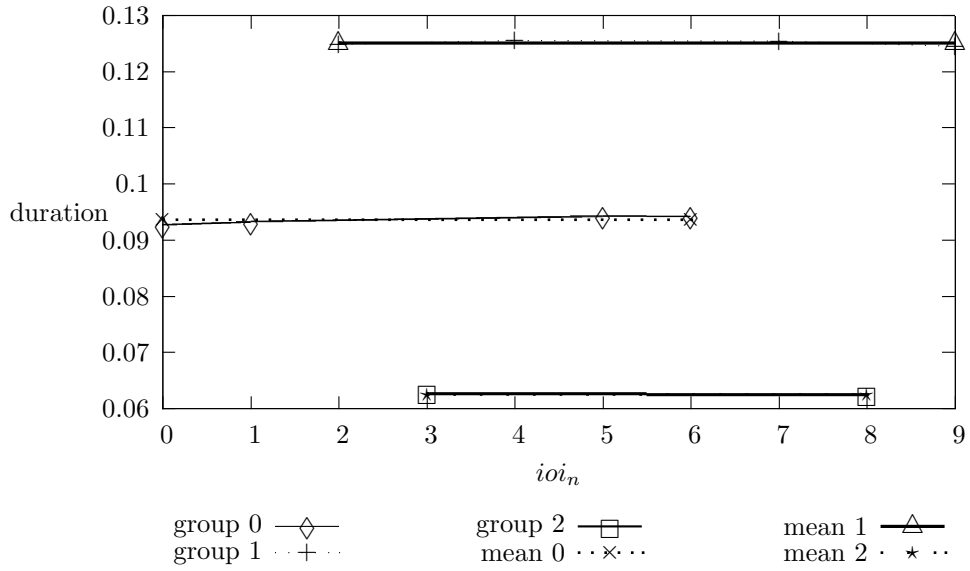


Figure 5-9: Bars 9-12 of the 3-2 *son clave* pattern presented earlier. An approximation of the mean values yields: $\{1/11, 1/8, 1/16\}$, which is very close to the original set of IOIs $\{1/12, 1/8, 1/16\}$.

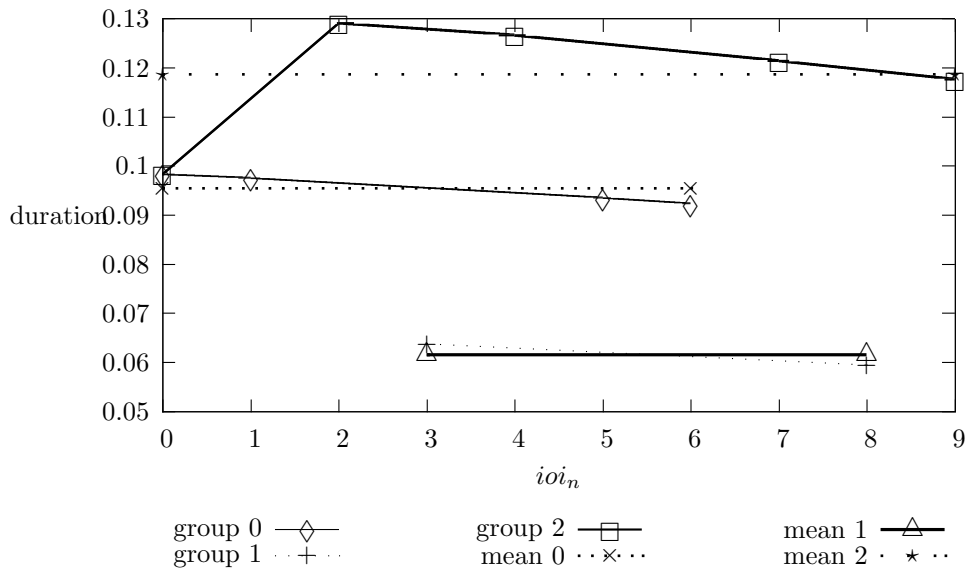


Figure 5-10: Bars 13-16 of the 3-2 *son clave* pattern presented earlier. An approximation of the mean values yields: $\{1/11, 1/9, 1/17\}$, which is very close to the original set of IOIs $\{1/12, 1/8, 1/16\}$.

but only if $r_i \leq 1$, then $r_i = r_i \times -1$. If the algorithm detects a change of sign from negative to positive r_i, then it will set a marker at d_i to start a new window. If there is a consecutive sequence of positive values for r_i, then the peak value for r_i in the sequence will trigger the new window marker.

To avoid the triggering of new window in case there is only a minute change from long to short duration, we introduced a minimum change threshold $\tau = 1.2$ which is equivalent to the change from a quintuplet to a sextuplet duration, i.e., 6/5.

In order to keep resulting windows from this process at a minimum length, a window length threshold can be set. We achieved successful quantisation outcomes with a minimum window length of 1.3 seconds, i.e., two times the indifference interval of 650 milliseconds, see section 2.4.

The windowed data sets of durations are then sent to our quantisation and transcription system. Another agent then applies the correct metrical levels to the quantised material and he controls the subsequent assembly of the quantised windows into a continuous transcription of the score. This agent is currently under development.

Figures 5-11 and 5-12 show an example of the computer-generated *son clave* rhythm that undergoes a continuous tempo change. Here we see that our duration ratio r peaks at significant changes from longer IOIs to shorter IOIs. If there is more than one consecutive positive value of r, the maximum positive value of that sub-sequence is chosen as the ultimate peak. The windowing algorithm then takes all durations between the index of the new peak that has been detected and the index of the previous peak. If there is no previous peak, as for example at

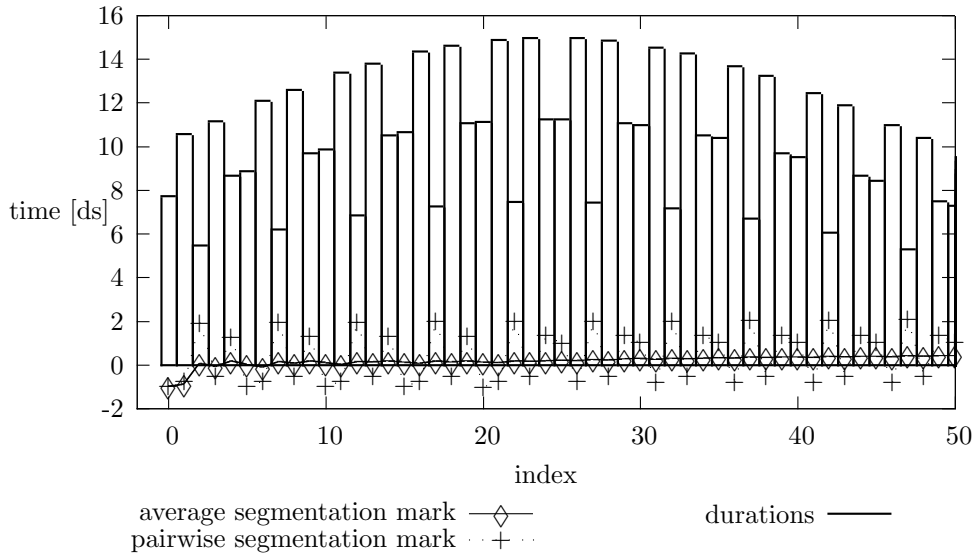


Figure 5-11: Onset segmentation of a tempo-varied stream of *son clave* rhythm, part 1

the beginning of a musical sequence, then all durations from the beginning to the current peak fall into the first window generated by this algorithm.

In our son clave example, it is noted that the cuts are always made at the same position in the repetitive rhythmic pattern although the tempo changes significantly and continuously.

5.4 Method of Overlapping Windows

The second method for windowing of onsets, separate from our motion index function (see previous section 5.3, p.127) consists in taking a small number of onsets per window and sliding the onsets window across the data one onset at a time evaluating each of those windows and comparing the quantised outcome of each window with the mean IOI value detected for every duration class within the same window. An evaluation of the indigestibility, see equation 3.7, of the quantised IOI ratios that will in different combinations represent the mean values of the duration classes is leading us to identify the best quantised IOI representation for each duration class. This evaluation will be described in detail under section 6.2. The size of the window is modelled after the magic number 7 ± 2 that plays a significant role in human cognition tasks, see Miller (1956). It describes the number of separate cognitive entities that can be kept simultaneously in short-term memory. The choice of $n = 7 \pm 2$ as the size of the sliding window yields reliable outcomes, which we demonstrate with the following example. Example figure 5-13 shows the IOIs of the tempo-modulated *son clave* rhythm we used earlier together with mean values of the duration classes that have been detected by our grouping algorithm. For each of the overlapping analysis windows, our grouping algorithm generates two or three duration classes, symbolised by 'o-o' in figure 5-13 and figure 5-14. As we will describe in the

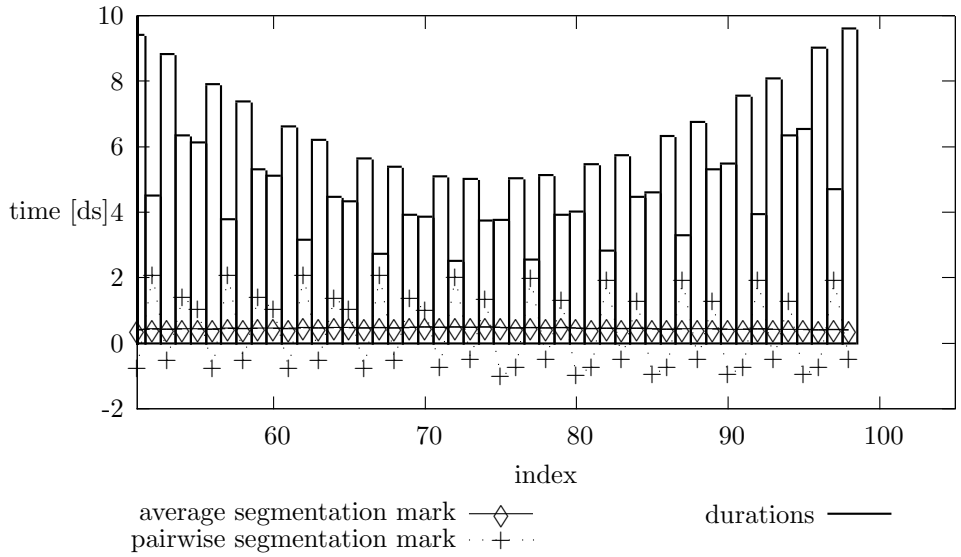


Figure 5-12: Onset segmentation of a tempo-varied stream of *son clave* rhythm, part 2

next chapter, a quantisation algorithm will search an area around each duration class for an integer ratio within a filtered Farey Sequence that is suitable for representing the duration class in CPN. In our example, this search algorithm has found the fractions

$$0.0625, 0.083333, 0.111111, 0.125, 0.166667$$

or:

$$\tfrac{1}{16}, \tfrac{1}{12}, \tfrac{1}{9}, \tfrac{1}{8}, \tfrac{1}{6} \quad ,$$

symbolised by '$|-|$' in figures 5-13 and 5-14. Combinations of three fractions are built and tested to see whether they would be close to the performed durations and close to a metrical grid. The details of this algorithm will be covered in the next chapter.

After the application of these cost functions, the quantiser arrives at the following set of fractions:

$$\tfrac{1}{12}, \tfrac{1}{9}, \tfrac{1}{6} \quad .$$

This set of fractions represents the three duration classes in correct proportions, which form part of the son-clave rhythm.

The quantiser and transcription process work in such a way as to give preference to those IOIs that show a low indigestibility value, which is based on the denominator of the fractional representation of the IOI.

The outcome demonstrated here is very important insofar as it represents a viable method for tempo detection. It further supports the quantisation part of our program because it eliminates the necessity for manual input of markers indicating downbeats as starting and end points of the analysis window. Figure 5-15 is a rhythm performed live with consecutive subdivision of beats into 2, 3 and 4 notes. Simple quantisers struggle normally with the proper distinction between 3- and 4-note subdivisions, e.g., eighth triplets versus 16th-notes. Within a sliding

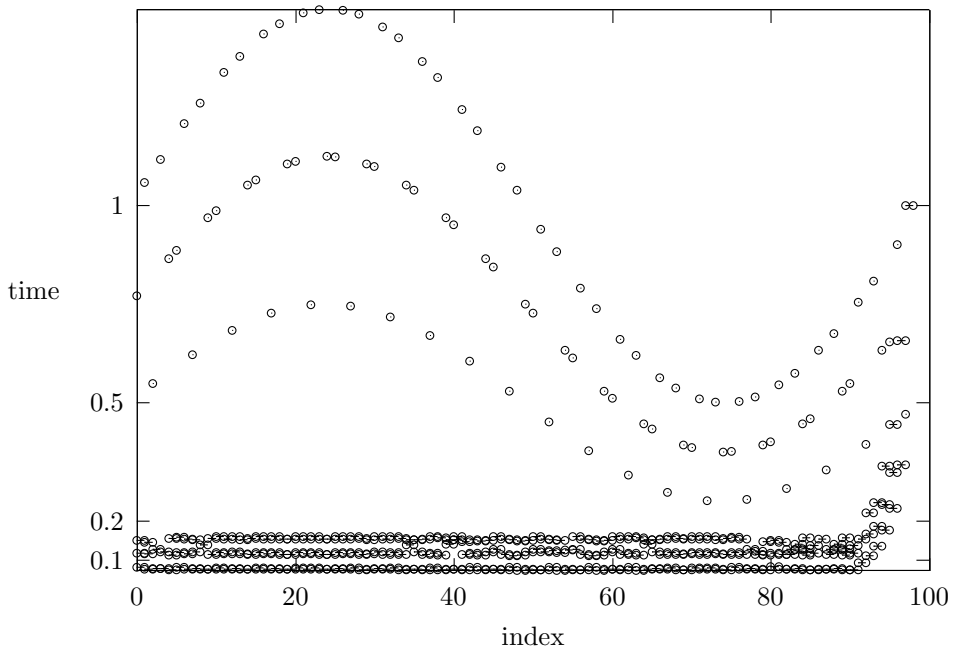


Figure 5-13: Tempo-modulated *son clave* track with duration classes detected between 0.0 and 0.2.

window, the durations are often not aligned to the underlying beat of the performance, for example a window could start at the second duration of a triplet. And, although this happens, our grouping algorithm remains precise in detecting the three duration classes involved.

5.5 Some Further Examples of Grouping

The following figures show the compound rhythms of the opening bars of Herbie Hancock's piece *Chameleon* from the record *Head Hunters*. Figure 5-16 shows the rhythm of the bass-line pattern, which is quite similar to the 3-2 *son clave* pattern. A characteristic difference is the three eighth-notes anacrusis, which creates an interesting tension because the way that the bass-line is being played on a synthesiser makes the perception of the downbeat quite ambiguous. This tension lasts until the drum-set starts, see figure 5-18, which resolves the riddle. Of course for the listener who knows the piece this effect deteriorates a bit. Our grouping algorithm detects all duration classes of the compound rhythms successfully as shown in figures 5-17 and 5-19.

5.6 Summary and Discussion

We have presented in this chapter one of our core processes within our system of automated quantisation and transcription, namely the automated grouping of IOIs into duration classes.

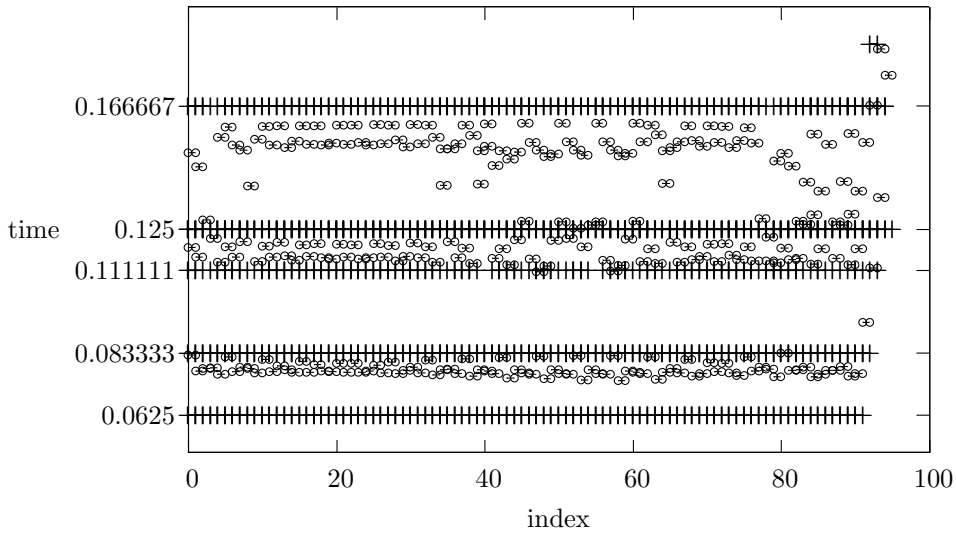


Figure 5-14: Zoomed-in graph of figure 5-13. Comparison of candidates for quantisation (|-|) and duration classes detected (o-o). The y-axis shows normalised IOIs.

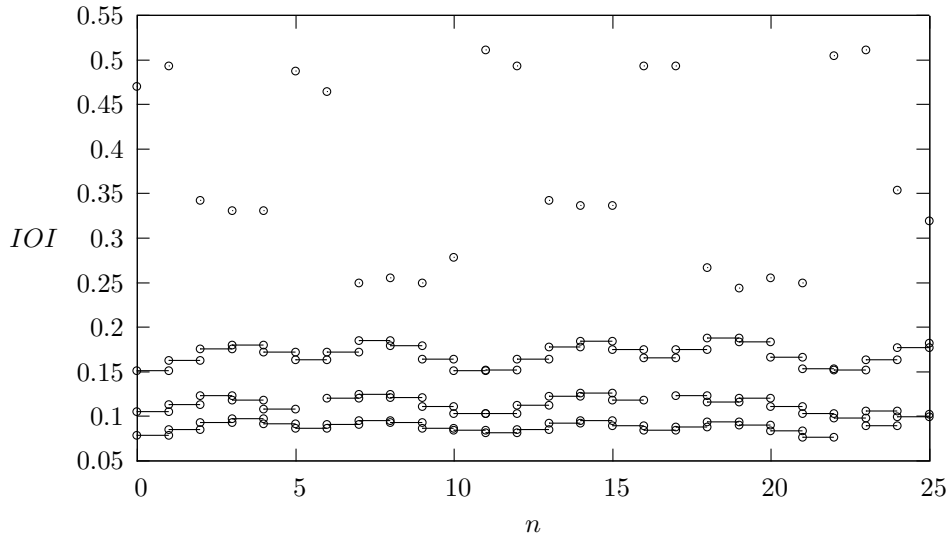


Figure 5-15: Live played rhythm pattern alternating subdivisions in 2, 3 and 4. IOIs (o) and duration classes detected (o-o) using a sliding window

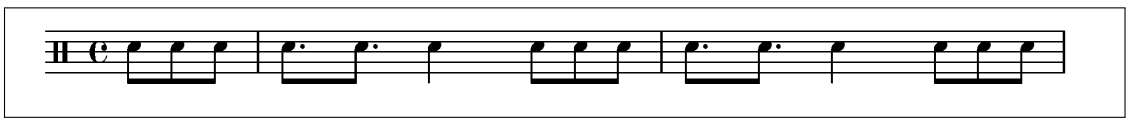


Figure 5-16: *Chameleon*: bass-line pattern

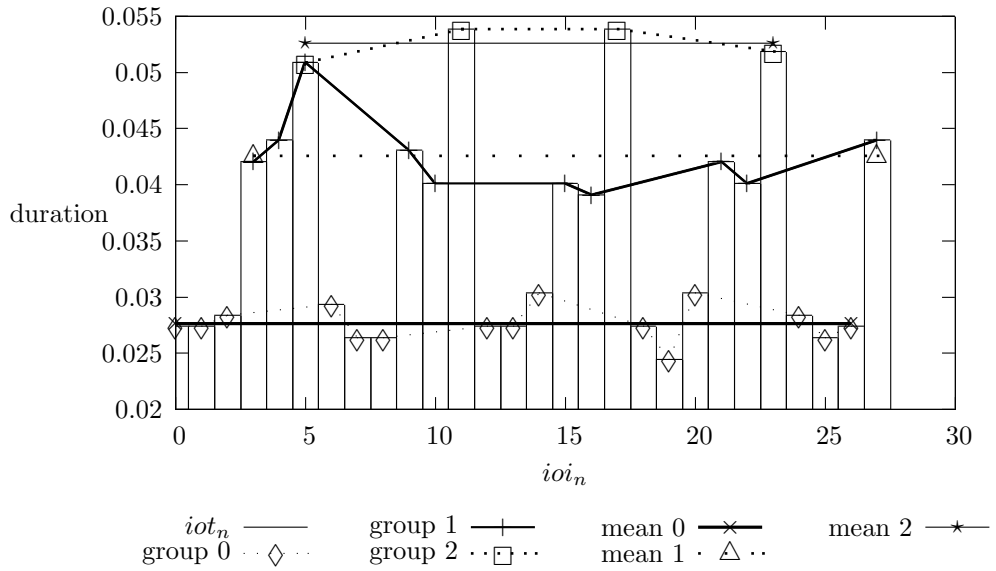


Figure 5-17: Groups detection for Herbie Hancock's *Chameleon*: Rhythm of the bass-line at the beginning. An approximation of the mean values yields: $1/37, 1/24, 1/19$, which is close to the original set of IOIs $1/36, 1/24, 1/18$ or $1/18, 1/12, 1/9$. Transcribed into a 4/4 bar, 1/18 represents the quaver, 1/12 a dotted quaver, and 1/9 represents the crotchet of the rhythmic pattern. Note that the variation of the IOIs strongly indicate a human player (Hancock) rather than a quantising sequencer program at work. The nPVI values for the manually played sequence is 23.1224, those of the groups 1-3 are: 6.59236, 7.40028 and 3.15313.

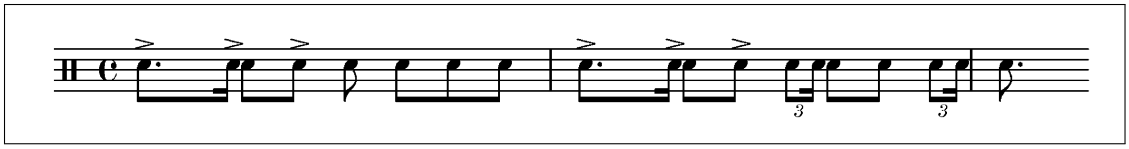


Figure 5-18: *Chameleon*: compound bass-line and drums pattern

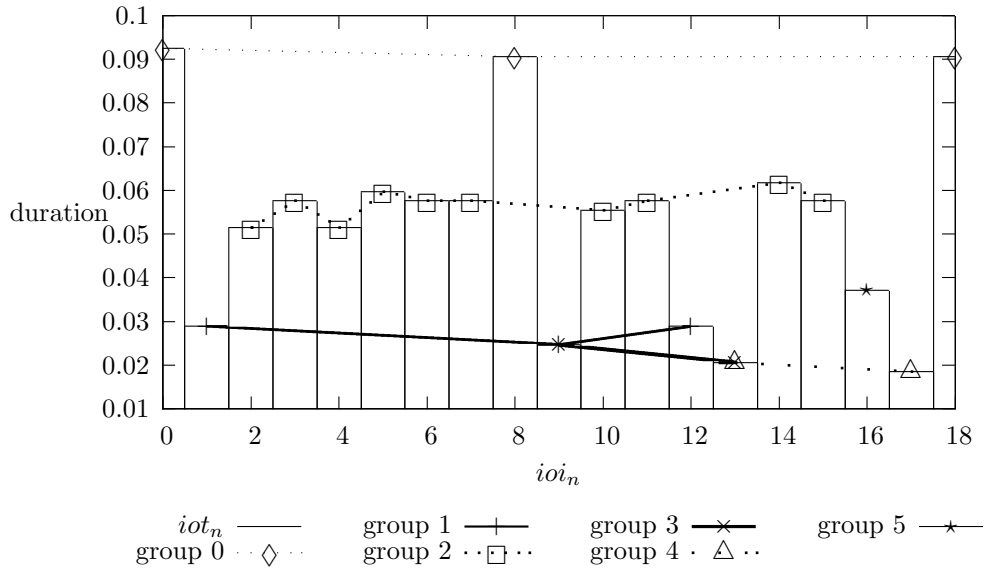


Figure 5-19: Groups detection for Herbie Hancock's *Chameleon*: Rhythm of the bass-line now joined by the drum set.

We have chosen a different design from traditional K-means clustering, mainly in order to avoid the random initialisation of clusters and its somewhat unpredictable influence on the outcome of the result, e.g., the question remained how many clusters should be initialised.

The grouping algorithm has been tested with both Gould recordings of the *Aria* from the *Goldberg Variations*. We found that, on average, approximately 90% of all note durations were correctly grouped into their corresponding duration class.

A continuous stream of discrete note onsets can be translated into IOIs, which are segmented into small windows of 7 ± 2 IOIs. Sliding this window across the buffered input data provides estimates of the duration classes present in the performance. This result is important for the research in tempo tracking and beat induction. The partly extreme tempo variations that can happen within the same duration class are correctly traced. The sliding window technique prevents the algorithm from the danger of mixing overlapping duration classes. Ambiguities in IOI groups, i.e., one or two durations shared between neighbouring groups, can easily be resolved by comparing the duration with both means of the groups in question. The group that has its mean closest to that duration value will claim it, and consequently the other one will loose it. This will then provide unique duration classes per each window of analysis.

In order to analyse an arbitrary stream of onset data from the performance of an unknown score we have developed a grouping algorithm. Its aim is to find duration classes within an analysis window. We found by experimentation that these windows cannot become infinitely large. They have to be bound by some perceptual timing limits. The challenge for our automated transcription system is to establish the boundaries for the analysis windows. We have experimented with three forms of windowing:

1. By placing markers by hand according to downbeats recognised by ear,
2. by using a sliding window analysis, and
3. by applying an automated segmentation algorithm based on the analysis of consecutive duration ratios.

Solutions 2 and 3 are informed by certain timing thresholds that have been reported by London (2004). A single window cannot grow above 6 seconds in length. It also features a minimal length of 1.3 seconds, which is twice the average indifference interval. The indifference interval is a perceptually important period of pulsation. We want to make sure that an analysis window is likely to span for at least two of these pulses. The threshold for the duration ratios in solution 3 is 1.2, which is the ratio between a sextuplet and a quintuplet (6:5). The aim for solution 3 is to separate windows when there is a change from long to short duration. We found experimentally, that one needs to prevent changes that are too small, i.e., a duration ratio < 1.2, to trigger the window segmentation, because such a small change is more likely due to an expressive bending of time that a performer applied to a sequence of equal durations. It is advantageous to separate windows at the onset of a beat or metrical pulse. Without relying on a dedicated beat tracking algorithm we found that the above thresholds, within the window segmentation process, improve results of the subsequent grouping and quantisation algorithms, because the chosen segmentation onset of the window is more likely to coincide with the onset of a metrical pulse.

We have also experimented with noise modulated sequences of equal durations in order to find the threshold where the grouping algorithm would find two separated duration classes instead of a single one. This threshold is at approximately 17 % noise modulation.

A separation of equal durations into separate duration classes can be reversed during the subsequent quantisation and transcription process. This is made possible because the mean values of the duration classes are centres of search ranges, which are allowed to overlap. The quantisation algorithm searches within a filtered Farey Sequence for relatively simple integer ratios. If two search ranges overlap significantly, it is possible that the same integer ratio for CPN will be assigned to two different duration classes. In this way, a former separation of performed durations on the basis of the grouping algorithm will be reversed by the quantiser.

Chapter 6

Farey Sequence Grid Quantisation

6.1 Introduction

After our grouping algorithm has identified the duration classes $\{E_i\}$ and their mean values $\{M(E_i)\}$ for each individual analysis window W, the next task is to quantise $\{M(E_i)\}$. Because a Farey Sequence can model musical rhythms and metrical hierarchies, see sections 3.2 and 3.3, the quantiser searches within a filtered Farey Sequence for appropriate integer ratios, which can be used for the notation of the performed note durations. A weighting function for integer ratios, which we derived from Barlow's indigestibility function, see equations 3.7 and 3.8, leads to the identification of quantised durations within in a filtered Farey Sequence that would match a particular value of $M(E_i)$ within a certain range of error. Per duration class $M(E_i)$ we build a pool of four close candidates and construct a set with all possible combinations between them. From those quantised durations, new onset positions for the performed rhythmic events can be deduced. By mapping the quantised durations back to the performed IOI positions, a pool of results is then calculated per analysis window W. This pool of results is analysed and sorted by the transcription agent and converted into CPN.

We are now introducing our method of rhythm quantisation that is based on a filtered Farey Sequence combined with a weighting measure for integer ratios.

6.2 The Quantisation Algorithm

The first task was to extract timing information from musical performances. This has been achieved using the tools and techniques outlined previously in section 2.3 on page 36. We created various files of onset data from audio recordings using the *aubio* library by Brossier (2006). In a series of tests we aligned windowed fragments of onsets to inter-beat intervals or lengths of one, two bars or even entire phrases. We also created several data sets from MIDI

files and manual tapping. The recorded onset markers are in seconds. Per analysis window the onsets are converted into normalised durations, i.e., the sum of all durations in a window is 1. This means that the performed durations are regarded as a normalised L1 norm vector, like a Farey Sequence.

A filtered Farey Sequence F_{200} represents non-equidistant grid points for quantisation. The filtering process uses k-smooth numbers; in our case we tested the algorithm with 3-smooth numbers, which is sufficient for a broad number of musical cases, as we have seen in section 3.3. The order 200 for the Farey Sequence grid creates a fine timing resolution for the analysis window of 5 per mil of the total length of the window. For example, it would enable a resolution of 100 milliseconds for a 20 seconds analysis window. The mean values of all duration classes $\{M(E_i)\}$ per analysis window are calculated. They are then processed by the quantisation algorithm. Centred around each mean, $M(E_i)$, we build a quasi Gaussian search window, in which we scan the filtered Farey Sequence in search for fitting rationals. The size R of this search window is determined by:

$$R = \begin{cases} M(E_i) \times 0.75, & \text{if } |E_i| = 1, \\ M(E_i) \times 0.5, & \text{if } |E_i| > 1. \end{cases} \tag{6.1}$$

Note that search windows around $M(E_i)$ can overlap. This is useful for those cases where different duration classes obtained from the grouping algorithm might have the same duration in CPN.

All integer ratios from the filtered Farey Sequence F_{200} that fall inside the search window are weighted according to Barlow's indigestibility function in equation 3.8. This weight is further scaled by the quasi Gaussian window function in order to take into account the distance of the rationals to the value of $M(E_i)$. Four integer ratios with the highest weight are then picked as possible candidates for the quantisation of the particular duration class. We found by experiment that taking a minimum of four candidates increases the success rate of the quantiser. Next, all possible sets of combinations between individual quantisation candidates are formed by picking one candidate per duration class.

The following pseudocode illustrates how it works, see Algorithm 1. First, in order to find the candidates for quantisation, the function takes the mean of a duration class as input, as well as the total range of the search R, see equation 6.1, which is centred around the mean. The integer ratios of the filtered Farey Sequence are elements of a double-linked list that is a member of the object class performing the quantisation. The data type of the list is a structure in which the weight can be stored together with the integer ratio. The table of the quasi Gaussian window used in Algorithm 1 is filled with values of the function

$$f(x) = exp(-x^2) \tag{6.2}$$

sampled at 1024 equidistant points in the range between $x = -1$ and $x = 1$, see figure 6-1. The purpose of this algorithm is to quantise the means of the duration classes. After the weighting function finished its work, the resulting list of weighted rationals is sorted from low to high

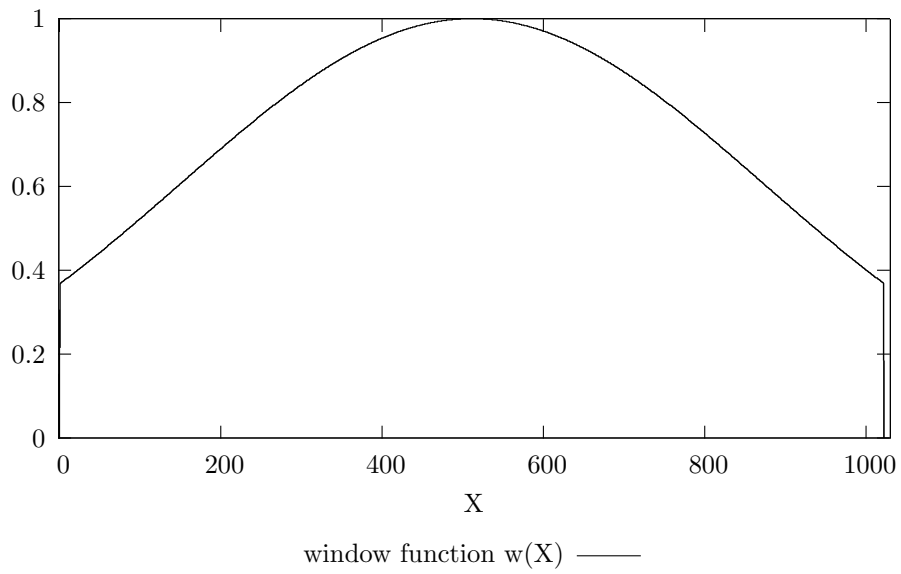


Figure 6-1: $f(x) = exp(-x^2)$ sampled in $[-1 \ldots 1]$

indigestibility. The first four elements from that list are then stored. This process is repeated for all means of duration classes, $\{M(E_i)\}$.

Algorithm 1 Attach weights to integer ratios of filtered F_{200}

```
{Input: mean, range, list} {Output: modified list}
max ⇐ (mean + 0.5 × range)
min ⇐ (mean − 0.5 × range)
length ⇐ length of Gaussian table
while not past the end of list do
  element ⇐ current element of list
  if element.value ≥ min ∧ element.value ≤ max then
    n ⇐ length × (element.value − min)/(max − min)
    w ⇐ value of Gaussian table at index n
    element.weight ⇐ w/(indigestibility of element.value.denominator)
  else
    element.weight ⇐ 0
  end if
end while
```

We then compute all possible combinations between the quadruplets of quantised means. The total amount of combinations we need to test is 4^d, with d = number of duration-classes, or $|\{M(E_i)\}|$. Each one of those combinations is a potentially viable set of quantised durations for the entire window W. Our aim is to find the set of durations that is not only as close as possible to the performed durations but also easy to read from a score in CPN. Therefore, the next task is to find out which combination of quantised durations is the best match for the performed durations. In order to achieve this, we found via experimentation that there are two

cost functions to be used.

The first cost function establishes a value for the metrical smoothness of the grid of onset points derived from the set of quantised durations of the analysis window. This is done in order to measure where exactly the quantised durations align with a metrical grid of onset points. Let Q be the set of new, normalised onset points within analysis window W that are derived from a particular combination of quantised durations. Q is a filtered Farey Sequence too, because the quantised durations were extracted from the filtered Farey Sequence F_{200}

$$Q = \{q_1, q_2, q_3, q_i, ...q_{|W|}\}$$

with $i = 1, 2, 3...|W|$. We then measure the indigestibility υ_i for each integer ratio q_i in the following way. Let a/b represent the integer ratio q_i, then we calculate:

$$\upsilon(q_i) = \xi(a) + \xi(b).$$

$\xi(n)$ is Barlow's indigestibility function, see equation 3.7. We arrive at the total value for the alignment of the quantised durations to a metrical grid of onset points by taking the sum:

$$\Upsilon(Q) = \sum_{i=1}^{|W|} \upsilon(q_i). \tag{6.3}$$

With this cost function we would like to achieve that the notation of the quantised durations uses relatively simple integer ratios in terms of the prime number composition of numerator and denominator. Those relatively simple integer ratios should also align with a metrical grid of onset points as close as possible. We keep in mind that, according to section 2.2, CPN exhibits a preference for simple integer ratios based on 3-smooth numbers, which yield relatively small indigestibilty values too. CPN is also based on a metrical grid of onset points by using a similar set of integer ratios, see section 2.8 about musical metre. A lower value of $\Upsilon(Q)$ indicates a closer distance to a metrical grid. We found through experimentation that with such a set Q one would arrive at a relatively simple notation.

The second cost function we take into account is the Euclidean distance measure. Out of the set of possible combinations of quantised durations, let Q be a normalised set of quantised durations:

$$Q = \{q_1, q_2, q_i, ...q_{|W|}\},$$

and let W be the normalised set of performed durations:

$$W = \{d_1, d_2, d_i, ..., d_{|W|}\},$$

then we use the Euclidean distance measure

$$d_E(Q, W) = \sqrt{\sum_{i=1}^{|W|}(q_i - d_i)^2} \tag{6.4}$$

With this distance measure we can find out how far a particular set of quantised durations deviates from the actual performance.

In addition, the quantiser also counts how many slurs between the note durations are necessary in order to transcribe the durations into CPN. Finally, we count how many tuplets are needed for the transcription of Q. By keeping the number of slurs and tuplets to a minimum one can support the clarity of the notated outcome, so that the quantised score should be relatively easy to read. The number of beats per window can be supplied by the user, or it can be calculated automatically by the transcription algorithm.

Per analysis window W we finish with a set $\{Q_i\}$, with $i = 1, 2, 3, ..., 4^d$, and with $d =$ number of duration-classes, or $d = |\{M(E_i)\}|$. We create a new set U containing up to sixteen possible solutions based on selected elements from $\{Q_i\}$. First, we take from $\{Q_i\}$ eight sets with smallest $\Upsilon(Q_i)$ after equation 6.3, and copy them into U. Then, we copy eight sets from $\{Q_i\}$ with smallest $d_E(Q_i, W)$ into U, according to equation 6.4. U will contain sixteen sets and is then sorted from low to high value for $d_E(Q_i, W)$. Finally, a solution Q_i is picked from U that exhibits the least numbers of slurs needed for the transcription. The algorithm gives further preference for a solution that features the lowest number of subdivisions in CPN. The number of slurs and subdivisions can be found by calculating a transcription for each solution in U, of course without invoking the rendering backend. We will now explain how this transcription algorithm works.

6.2.1 The Transcription Algorithm

We have used LilyPond for the rendering of the quantised durations, but also tried a combination of Fomus and LilyPond. Fomus has the advantage of also providing an export function for MusicXML, a file standard that can be read by commercial notation packages like Finale™ or Sibelius™. For the test examples given in section 6.3, only LilyPond has been used.

For the notation rendering via LilyPond an input file is generated algorithmically. It starts with the sets of normalised quantised durations, one for each analysis window. Per set we calculate various metrical grids and test how well they align with the onset points provided by the quantised durations.

The metrical grids are based on equidistant beats. The number of beats per grid are taken from this set of prime numbers:

$$B = \{2, 3, 5, 7, 11, 13, 17\}$$

In CPN, beats have a duration that is given as an integer ratio just like note durations. We

therefore form a set of integer ratios with the elements of B:

$$C = \{\frac{2}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \frac{11}{2}, \frac{13}{2}, \frac{17}{2}\}$$

As denominator one can chose a different power of 2, because in CPN we also encounter the denominators $4, 8, 16$ and 32. However, it is not so much a question of which denominator of the metre is chosen, since it does not much affect the metrical structure (the user can chose a different power of 2). What does matter is the number of beats of the underlying metrical grid that we are going to estimate here. This is encoded in the numerator of the time signature of the metre.

Let Q_i be a set of quantised and normalised durations of analysis window W:

$$Q_i = \{q_1, q_2, q_j, ...q_{|W|}\},$$

with $j = 1, 2, 3, ..., |W|$. We then calculate k scaled sets of durations per element c_k of C, with $k = 1, 2, 3, ..., |C|$:

$$Z_k = \{q_j \times c_k, j = 1, ...|W|\}$$

Then, for each of the rational numbers z_j in the set Z_k, we calculate Barlow's harmonicity function according to equation 3.10 and sum the results together.

$$H(Z_k) = \sum_{j=1}^{|W|} H(z_j)$$

The smallest sum of harmonicities $H(Z_k)$ marks the best fitting beat duration c_k, in the sense that a transcription of Z_k would require only very few slurs, if any, between note durations and that only those subdivisions of the beat are preferred in the transcription that have relatively low prime factors. This means that, for example, apart from the usual subdivision by powers of two and three, quintuplets, septuplets, 11-tuplets, etc., are less likely to appear in the transcription. Once the best beat duration is chosen, Z_k is fed into the subsequent transcription algorithm. Apart from the above process to infer the underlying beat pulsation, the algorithm employs techniques described in section 3.2.4, p.82, in order to transcribe the quantised IOIs into the format used by CPN, specifically for the formatting of a LilyPond input file. In addition to the above automated metre finding process for each of the analysis windows, the user can force a specific metre to be used in the transcription.

We will in the following section present a series of tests carried out with our algorithm. They are based on piano performances by Glenn Gould recorded in 1955 and 1981. We will also describe how we have solved the problem of detecting musical ornaments in a performance and how to transcribe them successfully into a usable notation.

6.3 Test Results

Here we present a series of test results that demonstrates the outcome of our grouping, quantiser and transcription algorithms. In this series of tests the knowledge of the downbeats is given to the program, because the onset data of the performance have been divided into bars by hand.

The following figures 6-2 to 6-10 display the original score of the *Aria* with Bach's ornaments written out in the way that Gould has played them. On the third staff, below the piano staves, we show the results of our quantiser. The quantised durations relate to the score as the compound rhythms played by both hands during the performance.

Optimum results were experimentally achieved when allowing a maximum digestibility of the tuplets of $C(1/7)$, see equation 3.8, p.107, and by allowing a maximum number of two slurs, i.e., ties between notes, per bar.

We have marked all positions of the ornaments by Bach in red, and the ones added by Gould in blue next to the relevant voices of the piano score. When it comes to the execution of the ornaments, there are notable differences between Gould's version from 1955 and his last recording in 1981. The reader may compare the last figure in the right hand in bar 6, the trills in the left hand in bars 8, 10, 21 and 22. Furthermore, in 1981, there is an appogiatura added in bar 11, the last note in the left hand. Another appogiatura occurs in 1981 in bar 19, left hand, just before the third beat. The following trill in the central voice is also played differently. Finally, there is a note inègale in the right hand in bar 26 that occurs only in 1981. As we have pointed out in section 2.6, p.46, ornamentations are open to variations by the individual player with regard to the timing and rhythmic structure of the particular ornament. Therefore, ornaments are a particularly challenging task for any quantisation and transcription program.

Written beneath the quantisation staff is the amount of indigestibility per quantised bar, $\Upsilon(Q_i)$, and the Euclidean distance between the set of quantised durations and the set of performed durations, $d_E(Q_i, W)$, for every bar. Both sets, the quantisation and the performance data, are always normalised, they are in fact normalised L1 norm vectors. Hence we conjecture that, in theory, the maximum Euclidean distance between two of these vectors can only reach an open upper limit of $\sqrt{2}$. However, it is unlikely that values close to $\sqrt{2}$ appear, because it would mean that the performance of the score would feature one bar with many close-to-zero durations and a single close-to-one duration. This is musically a rare case, which does not occur in our Bach example. Moreover, in order to maximise $d_E(Q_i, W)$, the grouping algorithm would have to swap a single long duration with a short one, which, according to the design of the algorithm in chapter 5, should never happen[1], because a single maximum duration would lead to its own duration class, and all close-to-zero durations would be grouped into a different duration class. In theory, however, only a swapping of position of the long duration would lead to a Euclidean distance of approx. $\sqrt{2}$ against the performance. In our experiments we have never seen such a large distance. Due to time constraints we have to leave the conjecture open that $d_E(Q_i, W) < \sqrt{2}$ is true. The experimental data measured so far do not show a higher limit.

[1] It would be a bug in the implementation if it occurred.

We also compared the performed durations per bar with the original score and calculated the Euclidean distance $d_E(W_l, J_l)$, with J_l representing the normalised set of compound note durations in bar $l = \{1, 2, 3, ..., 32\}$, as written in the score and including a metrical notation of the way in which Gould has played Bach's ornamentation. The largest distance one can detect between Gould's performance in 1955 and the score is in bar 22, with $d_E(W_{22}, J_{22}) = 0.207624$. In Gould's 1981 performance, the maximum Euclidean distance is in bar 21: $d_E(W_{21}, J_{21}) = 0.165846$. Tables 6.1 and 6.2 give an overview of the Euclidean distances between Gould's performance and the original score in the 4th column. The next column shows the distances between the quantised bar and the original score. Here, a zero distance indicates a perfect match between them.

6.3.1 The 1955 recording

For the 1955 recording, see table 6.1, there are 8 out of 14 cases, where there is a positive Euclidean distance measured between the quantised result and the original score. In these 8 cases, the distance is less than (or equal to) the distance between the performance and the score. The average distance per bar between quantisation and score is 0.023491, the average distance per bar between performance and score is 0.042481. When there is a distance between the quantised bar and the score, it shows the influence of the performer's expressive timing, because the minimum Euclidean distance between the quantised set of durations and the performed set of durations is one of the criteria during the search for the optimum result. For example, the quantisation of bar 5 (1955) shows that the quaver during the 3rd beat is played slightly longer than it is written on the expense of the following shortened semiquaver. One can interpret this result as a quantised version of Gould's expressive timing at that particular moment. A possible motivation for a musician to play it that way is the fact that one actually deals with a dotted quaver and a superimposed mordant, which is the shorthand notation normally used by Bach and other composers of the era. It is common practice in Baroque music to play single-dotted notes longer by taking time away from the following shorter note (Efrati, 1979; Fabian, 2003), see also section 2.6. This effect of over-dotted notes can be enhanced through an ornament placed on the long note, which is happening in bar 5. The resulting durations of the performance are then affected in such a way that it influences the quantisation of the performance. It has a positive effect in the sense that one is able to see in the quantised text where exactly expressive timing has a strong influence on the durations that are actually played. This might be useful for musicians and musicologists who want to study how Gould performed the music by Bach.

In bar 14, beat number 3, there is a similar case of a prolonged note followed by a shortened demisemiquaver, the c-sharp in the right hand. In bar 13 we find a case, where the quantisation results in a notated ritardando of the first semiquavers on beat 1. The grouping algorithm has collected the last semiquaver of that figure and the last semiquaver of the bar into the same duration class, which is quantised into the longer quintuplet quaver as opposed to an ordinary semiquaver. We interpret this result as an approximation of the original score that is informed

by the expressive timing of the performer. This result can also be observed in bar 6, where Gould plays the quintuplet with a little rubato, a slight accelerando over the five notes of the *cadence*[2] in the right hand, last quaver of bar 6.

Several cases of Gould's rubato playing shine through the quantised result. A short lingering on notes can also be observed in bars 7, 8, 9, 10, 11, 16, 17, 19, 22 and 32. Most of these cases do not affect the clarity of the quantised notation and are very informative about Gould's use of agogics. However, the expressive timing in bars 7, 11, 17 and 22 is not straightforward to transcribe. Complicated ornamentations are featured in bars 11 and 17. There are long trills and trills in both hands simultaneously, that make the task of quantisation and transcription particularly challenging. Bars 7 and 22 have both groups of over-dotted notes where the duration of the dotted notes are prolonged.

The Aria is divided into two halves after bar 16. Bar 16 has a harmonic cadence, a formal ending of the first half of the piece where musicians tend to slow down in order to indicate that the music of first 16 bars comes to a preliminary halt. The quantisation of bar 16 captures this nicely with a prolonged 'd' transcribed as a quaver during the first beat and a ritardando indicated by the following durations, i.e., the long 'd' on beat 1 steals a demisemiquaver from the following gesture.

From bar 26 onwards until the end of the *Aria* we find a continuous stream of semiquavers, which has been perfectly quantised. The last bar contains a final ritardando. This bar is special because the harmonic cadence ends with an appogiatura on the weak beat 3. The quantisation follows Gould's ritardando, which gradually slows down by factors of 2. In order to finish within one bar, the transcription of this bar starts with a demisemiquaver.

6.3.2 The 1981 recording

For the quantisation of this recording, the average Euclidean distance per bar between quantisation and score is 0.014751, and the average distance per bar between performance and score is 0.033573. The quantisation and transcription matches the original score perfectly in 24 bars out of 32, a 75% match. Bars 11 and 17 are still very challenging, as in the 1955 recording. In bar 11, the algorithm faces a long trill with an additional arpeggio figure (right hand, beat 1) and later an appogiatura in the left hand, which is the last demisemiquaver. Bar 11 is still rhythmically convincing, with a little rubato lingering on the first note of the original sextuplet trill figure, which shifts the subsequent durations proportionally. Still, it is possible to follow the intended accelerando during the long trill of the right hand. Bar 17 features two trills simultaneously in both hands on beat 2. There is additional rubato by Gould in the timing of the ascending figure on beat 3 in the right hand. Because of this detail, there is a similarity between the two recordings when it comes to the rubato timing of this bar.

The transcription of bar 13 indicates a slight ritardando during beat 2, after which the tempo resumes with the last two notes of the right hand. Bar 14 features the over-dotted note phenomenon on beat three: the c-sharp of the right hand is prolonged a little bit on the expense

[2]cadence is in this case the technical term for a particular type of ornament.

of the following d.

The ritardando in bar 16 has been successfully quantised, whereas in bar 32 one encounters a ritardando in proportional steps, similar to the 1955 recording.

Analysis of the quantisation of Gould's 1955 performance.					
bar	number of onsets	bar length in seconds	Euclidean distance betw. *Gould's performance* and original score	Euclidean distance betw. *quantised performance* and original score	bar
1	6	2.7575	0.0674779	0	1
2	6	2.8095	0.0339418	0	2
3	17	2.4205	0.0362363	0	3
4	10	2.9295	0.026371	0	4
5	6	2.926	0.0424895	0.020833	5
6	11	3.036	0.0341105	0.00937254	6
7	13	2.9615	0.0331713	0.0489802	7
8	10	3.872	0.0709284	0.102062	8
9	6	3.106	0.0420389	0.0340188	9
10	13	3.207	0.0575668	0.060875	10
11	21	2.9875	0.0230705	0.0496243	11
12	8	3.329	0.0455098	0	12
13	9	3.5205	0.0213791	0.0720083	13
14	10	3.6545	0.0264608	0.0147317	14
15	9	3.7325	0.0302118	0	15
16	9	4.2545	0.034634	0.0510314	16
17	20	3.2245	0.115762	0.102242	17
18	8	3.6135	0.0313371	0	18
19	15	3.4015	0.0370465	0.0170234	19
20	8	3.524	0.0226643	0	20
21	17	3.457	0.0336027	0	21
22	14	3.552	0.207624	0.085582	22
23	10	3.901	0.0280434	0	23
24	10	4.069	0.022005	0	24
25	8	3.3905	0.0186024	0	25
26	14	3.401	0.0373512	0	26
27	12	3.4795	0.0160471	0	27
28	12	3.503	0.0169429	0	28
29	12	3.3875	0.0285099	0	29
30	12	3.485	0.0162152	0	30
31	12	3.835	0.0281443	0	31
32	10	6.232	0.0739067	0.0833332	32

Table 6.1: Euclidean distances between the quantised durations of Gould's 1955 performance and the original durations of Bach's score. Values are measured for each bar together with the number of durations per bar and the length of each bar in Gould's performance.

Analysis of the quantisation of Gould's 1981 performance.					
bar	number of onsets	bar length in seconds	Euclidean distance betw. *Gould's performance* and original score	Euclidean distance betw. *quantised performance* and original score	bar
1	6	4.831	0.0243751	0	1
2	6	5.324	0.0224464	0	2
3	17	4.72	0.047737	0	3
4	10	5.584	0.0285121	0	4
5	6	5.584	0.0110546	0	5
6	10	5.503	0.0230716	0	6
7	13	5.488	0.0199194	0	7
8	11	6.177	0.0202216	0.0731173	8
9	6	5.869	0.0271574	0	9
10	11	5.875	0.0238282	0	10
11	22	5.474	0.0293574	0.0354104	11
12	8	6.169	0.020688	0	12
13	9	5.897	0.0867668	0.0351296	13
14	10	6.084	0.0867668	0.0196435	14
15	9	6.257	0.0186715	0	15
16	9	6.844	0.025393	0	16
17	20	4.876	0.0244918	0.0809852	17
18	8	5.568	0.0211017	0	18
19	16	5.323	0.0957753	0.0837659	19
20	8	5.475	0.013019	0	20
21	16	5.253	0.165846	0.0294628	21
22	13	5.444	0.0214225	0	22
23	10	5.491	0.0112145	0	23
24	10	6.256	0.0259337	0	24
25	8	6.293	0.022554	0	25
26	15	5.52	0.0143151	0	26
27	12	5.549	0.00669801	0	27
28	12	5.497	0.0125099	0	28
29	12	5.544	0.00899706	0	29
30	12	5.712	0.0126862	0	30
31	12	6.067	0.0150426	0	31
32	10	10.275	0.0867668	0.11453	32

Table 6.2: Euclidean distances between the normalised quantised durations of Gould's 1981 performance and the normalised original durations of Bach's score. Values are measured for each bar together with the number of durations per bar and the length of each bar in Gould's performance.

6.3.3 Early Test Results

Table 6.3 lists our early results of the quantisation[3]. The optimisation in series #2 shown in table 6.3 was achieved by recalculating only the bars that have failed in series #1. Those are the bars that were split into single beats before quantisation: bars 17, 19 and 32 for 1981, bars 7, 8, 11, 13, 17, 21, 22 and 32 for 1955. Note that these are early test results.

[3]Results are also available online at http://dream.cs.bath.ac.uk/transcriptions/

Recording details:		
year	1955	1981
onsets played (both hands)	359	351
ornaments played	48	49
Series #1 - prior knowledge of downbeat locations:		
onsets quantised correctly	219 (61%)	308 (88%)
ornaments transcribed	31 (65%)	41 (84%)
Series #2 - prior knowledge of single beat locations:		
onsets quantised correctly	352 (98%)	351 (100%)
ornaments transcribed	46 (95%)	49 (100%)
number of critical bars	8	3

Table 6.3: Early results of the quantisation of the Bach *Aria*

The ritardando at the end of the first part was transcribed successfully for both recordings (series #1). At the end of the *Aria* (played without repetitions) the last note with appogiatura is held with fermata in both recordings, however, this bar has been transcribed successfully in series #2 for both recordings.

Apart from the downbeat locations, the program had no prior knowledge of the score in these early tests. The onset calculation was based on an analysis window that hopped over the onset data and was aligned to the exact length of one bar. In some cases, shown as series #2, the size of the window needed to be reduced to the length of a beat. Therefore, we have devloped two procedures for automated window segmentation, see section 5.3 and 5.4, in order to eliminate the need to determine the window size by hand. The setting of window markers by hand is a useful feature for interactive quantisation processes, for example when composers work with with applications in the domain of computer-assisted composition (CAC). Today we have the possibility to evaluate the quantisation results also in terms of their score rendering qualities, i.e., their complexity as expressed by the maximum tuplet number and by the number of ties between notes per analysis window. In addition, we now take into account the Euclidean distance between the quantised durations and the performed durations per analysed window as one of the search criteria in order to maintain a close connection between the quantised transcription and the musically performed rhythms.

6.4 Summary and Discussion

The method we presented is successfully able to quantise onsets from performance data extracted from audio recordings, to use analysis windows aligned to one bar each, and to transcribe the results into CPN. The ornaments and combined rhythms of right and left hands were, with few exceptions, successfully transcribed from two different piano recordings of Bach's *Aria* of the *Goldberg Variations*.

The proposed quantisation algorithm shows a number of advantages. The grouping method is successfully tracking duration classes in human expressive performances. The approach of

using a sliding analysis window is in support of the quantiser because tempo changes can be followed accurately. The tempo-modified *son clave* rhythm was previously reported to be a hard problem for tempo tracking systems because of its syncopations, see for example Cemgil and Kappen (2003), yet the sliding window analysis as well as the method using our window segmentation algorithm is fully capable of returning a correct quantisation of the rhythm by taking into account major continuous tempo changes. The quantisation method is also robust in situations of a noisy signal that could originate for example from technically inaccurate playing or from the biological motor functions of a player. The quantiser can cope with varying degrees of rhythmic complexity via the score transcription process. The windowed analysis is computationally also very efficient and capable of real-time execution. A large part of the computational complexity is given by the number of duration classes d detected and the number of possible candidates c chosen per class to represent its duration: c^d. Because of a relatively small window of 7 ± 2 onsets, the number of duration classes is very likely to be fairly small, between 2 and 4. The number of candidates is constrained to 5, 4 or even 3 when d is lower, equal or higher than 6 respectively. Therefore the computation is not in danger of spending too much time on calculating all possible permutations of the quantisation result. Barlow's indigestibility function is highly instrumental when choosing the best candidates for the quantisation. The harmonicity function is very useful for finding the best beat-grid for each analysis window in an automated way and thereby finding the least complex notation. However, the user is able to impose a specific metre to be used for the transcription task. Both, harmonicity and indigestibility functions are being used for the first time successfully in order to address the issue of quantisation and transcription of musical performances. The system also performs well when quantising and transcribing performances of musical ornaments, like trills and mordants.

The particular challenges of the Bach *Aria* are caused by its many simultaneously executed ornaments, which are occasionally combined with syncopated rhythms. In addition, we are dealing with polyphonic voices between two hands, which produces sometimes complicated compound rhythms. For our main tests, the separation of the analysis windows was done by hand, which introduces the knowledge of the downbeat onsets into the program.

We have measured also the Euclidean distance between the set of quantised durations and the set of performed durations. We have found perfect matches between the score and the quantised result. 56 % of the bars in the 1955 recording, and 75 % of the bars from 1981 are perfectly matched. In addition, we obtain quantised results, which notate very clearly the tempo modifications introduced by the performer. The results suggest in these cases that it is possible to obtain a quantised and therefore readable version of the expressive changes of note durations. This is made possible, because, along with the measurement of indigestibility, one of the search criteria for the solution was the Euclidean distance between the sets of quantised and performed durations.

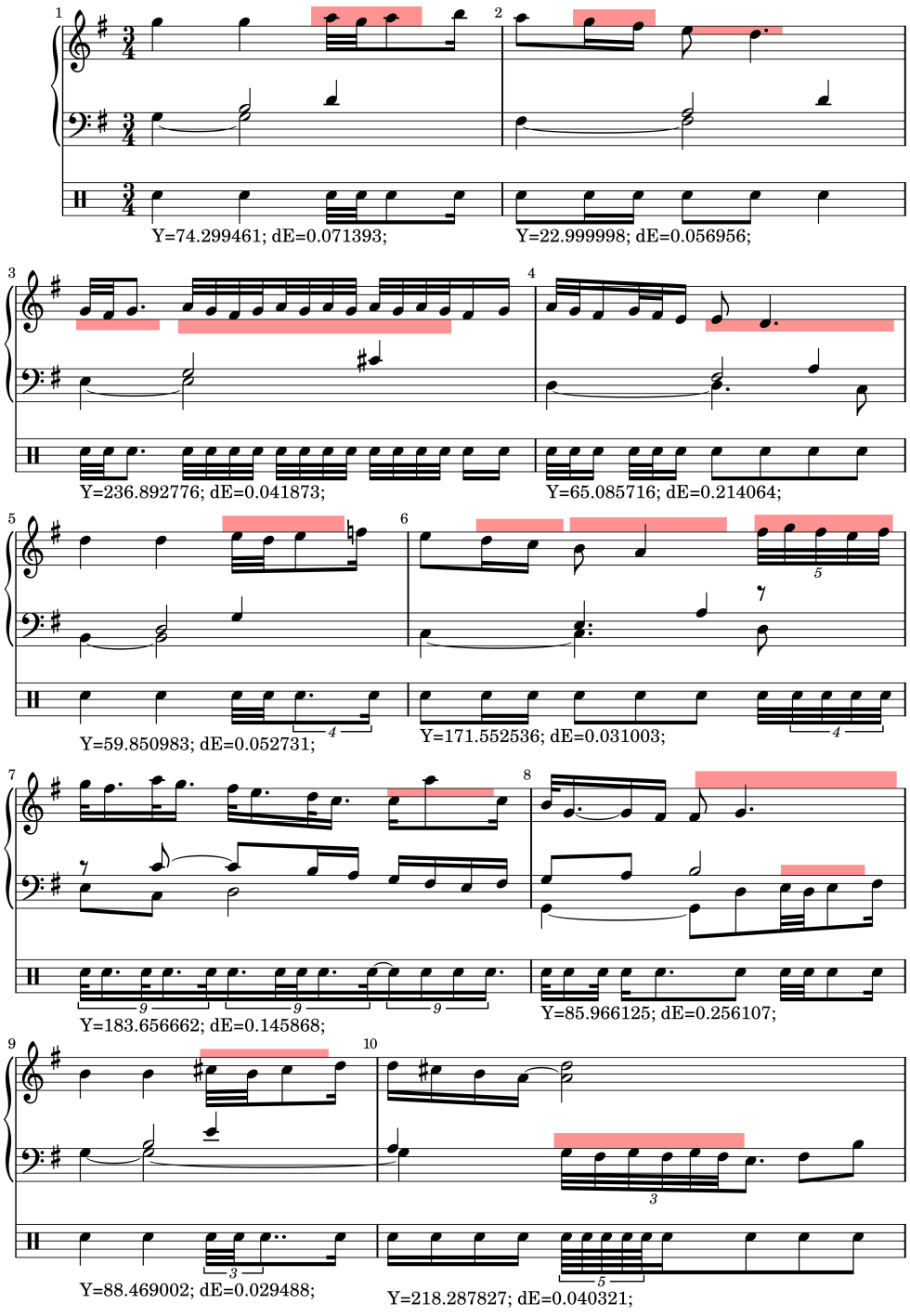


Figure 6-2: Quantisation of Gould's recording from 1955. Red colour indicates the position of Bach's ornaments. Blue colour denotes Gould's own ornamentation.

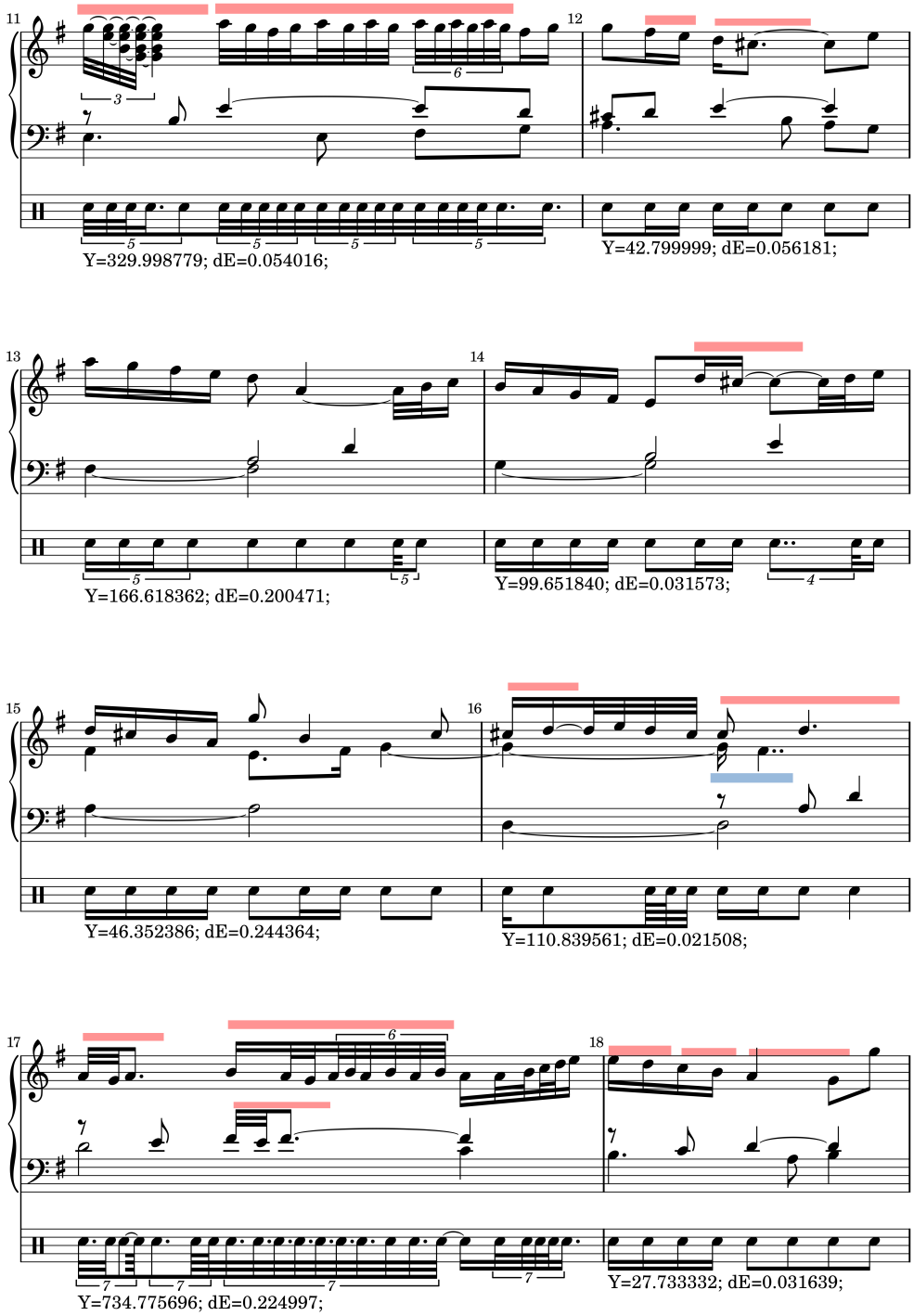


Figure 6-3: Quantisation of Gould's recording from 1955.

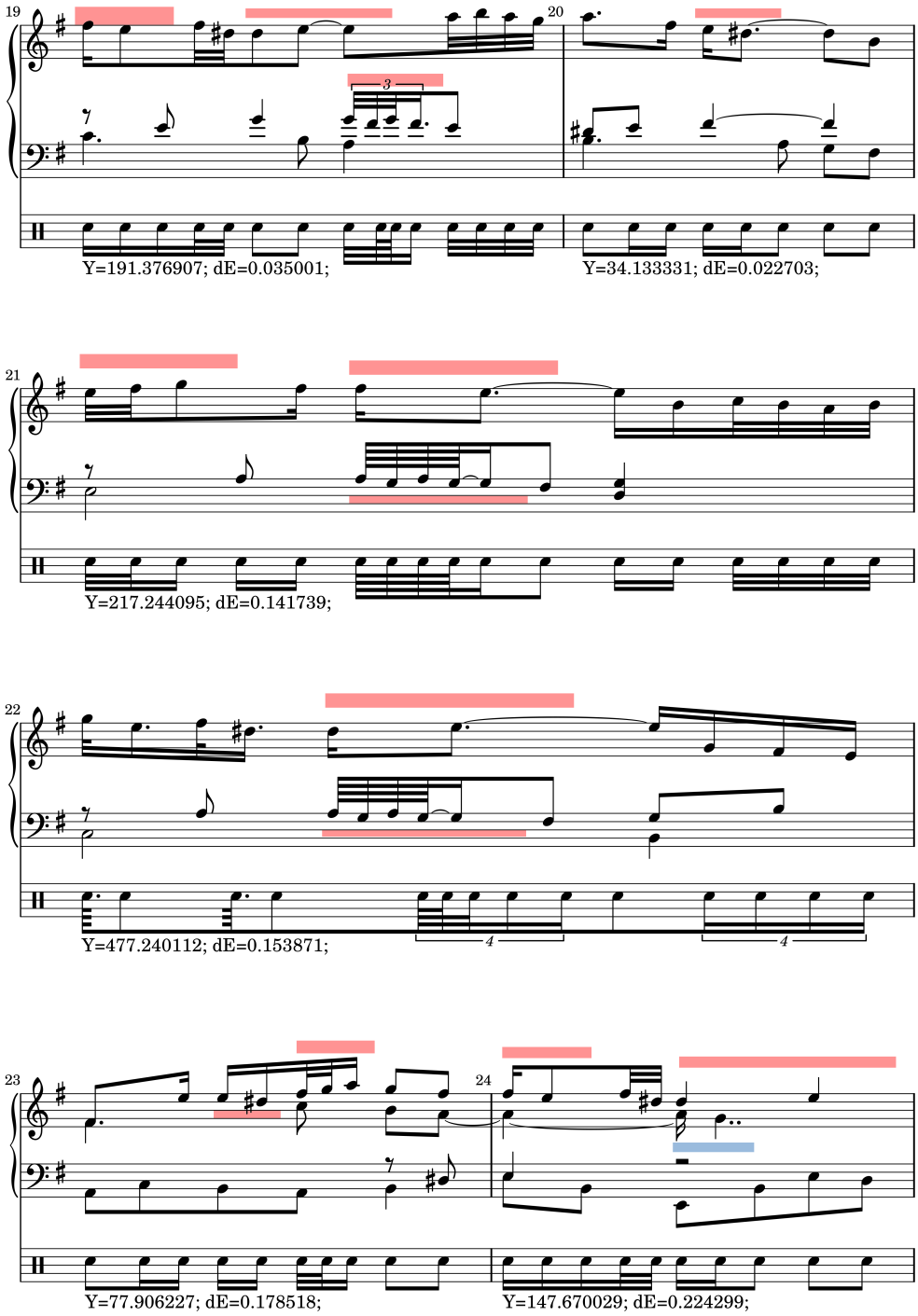


Figure 6-4: Quantisation of Gould's recording from 1955.

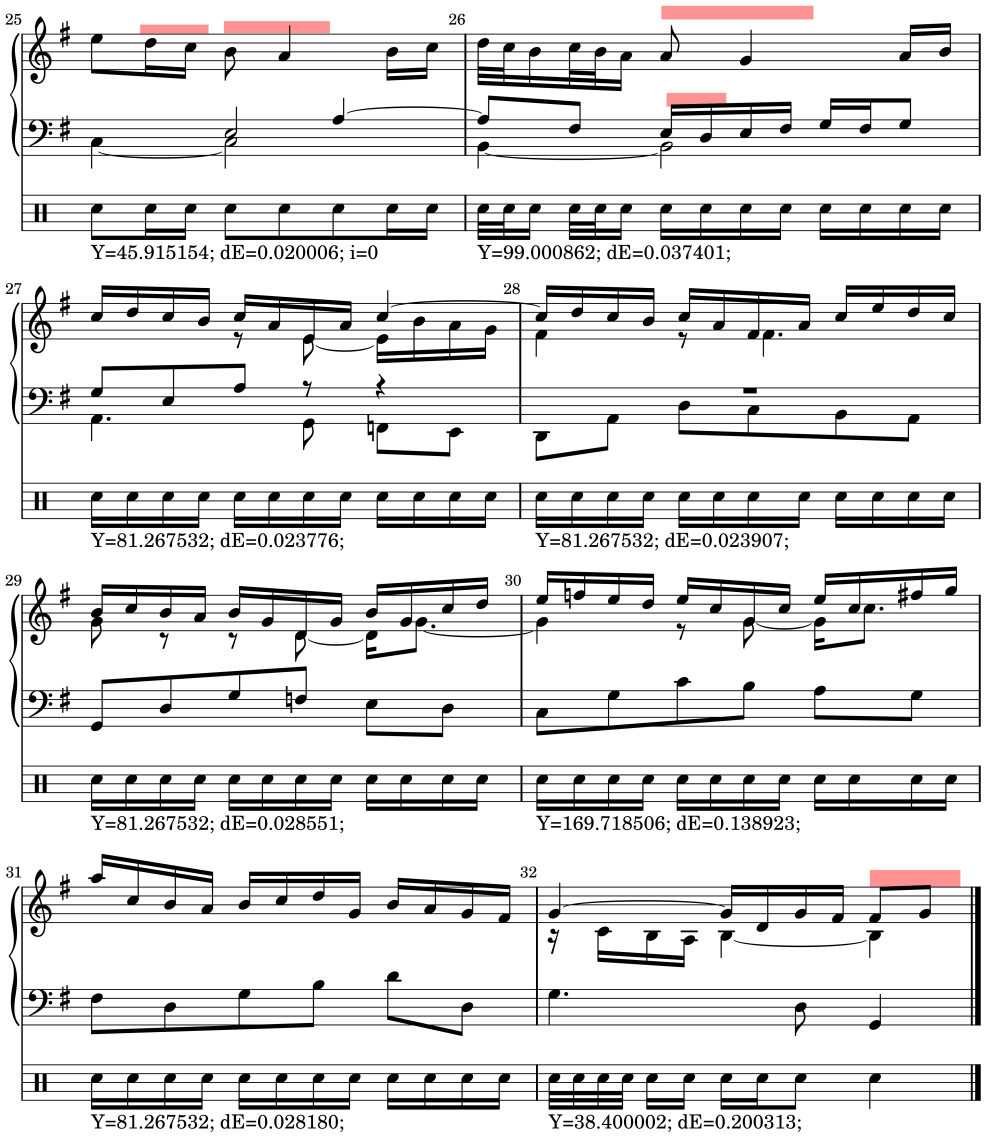


Figure 6-5: Quantisation of Gould's recording from 1955.

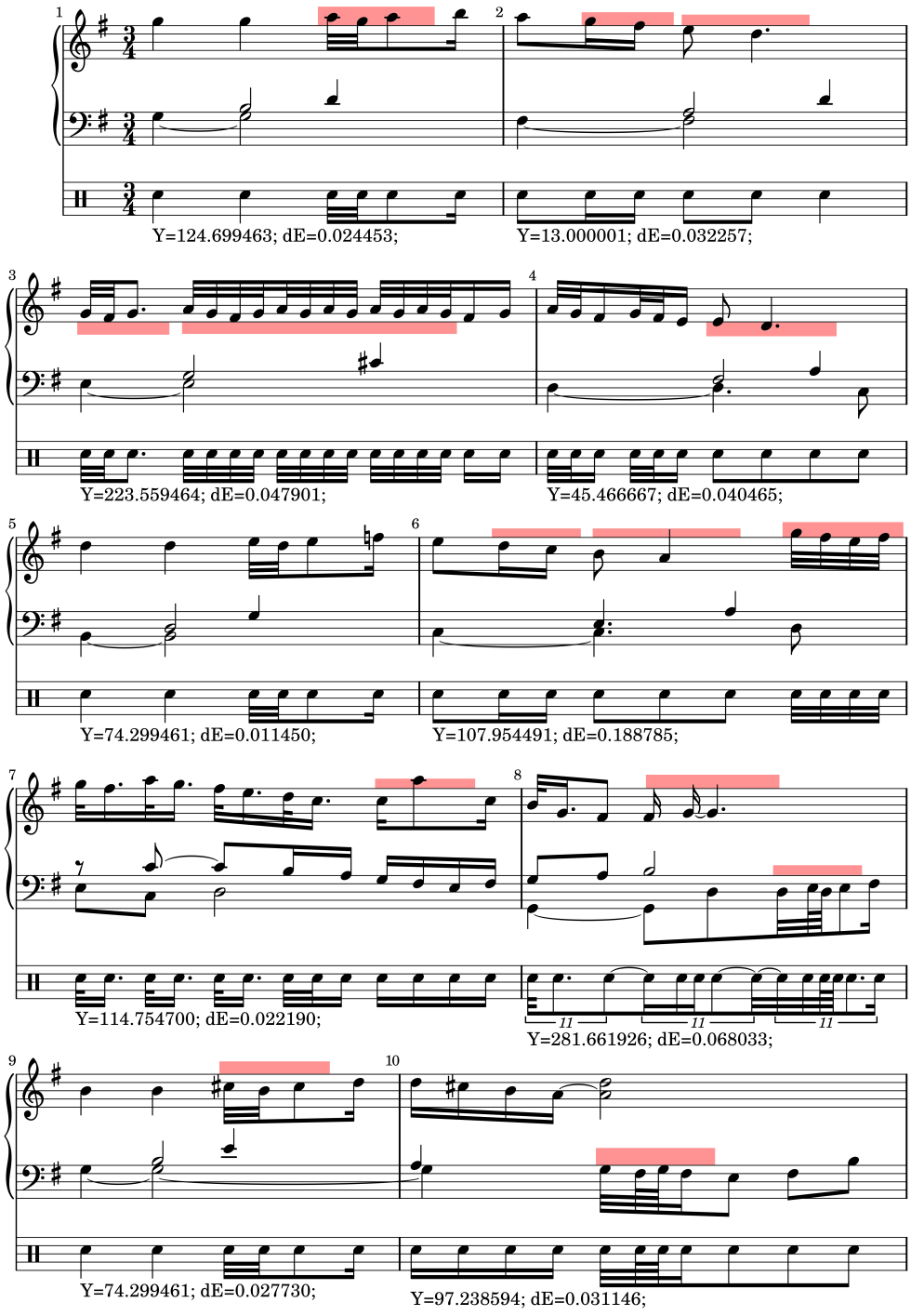


Figure 6-6: Quantisation of Gould's recording from 1981. Red colour indicates the position of Bach's ornaments. Blue colour denotes Gould's own ornamentation.

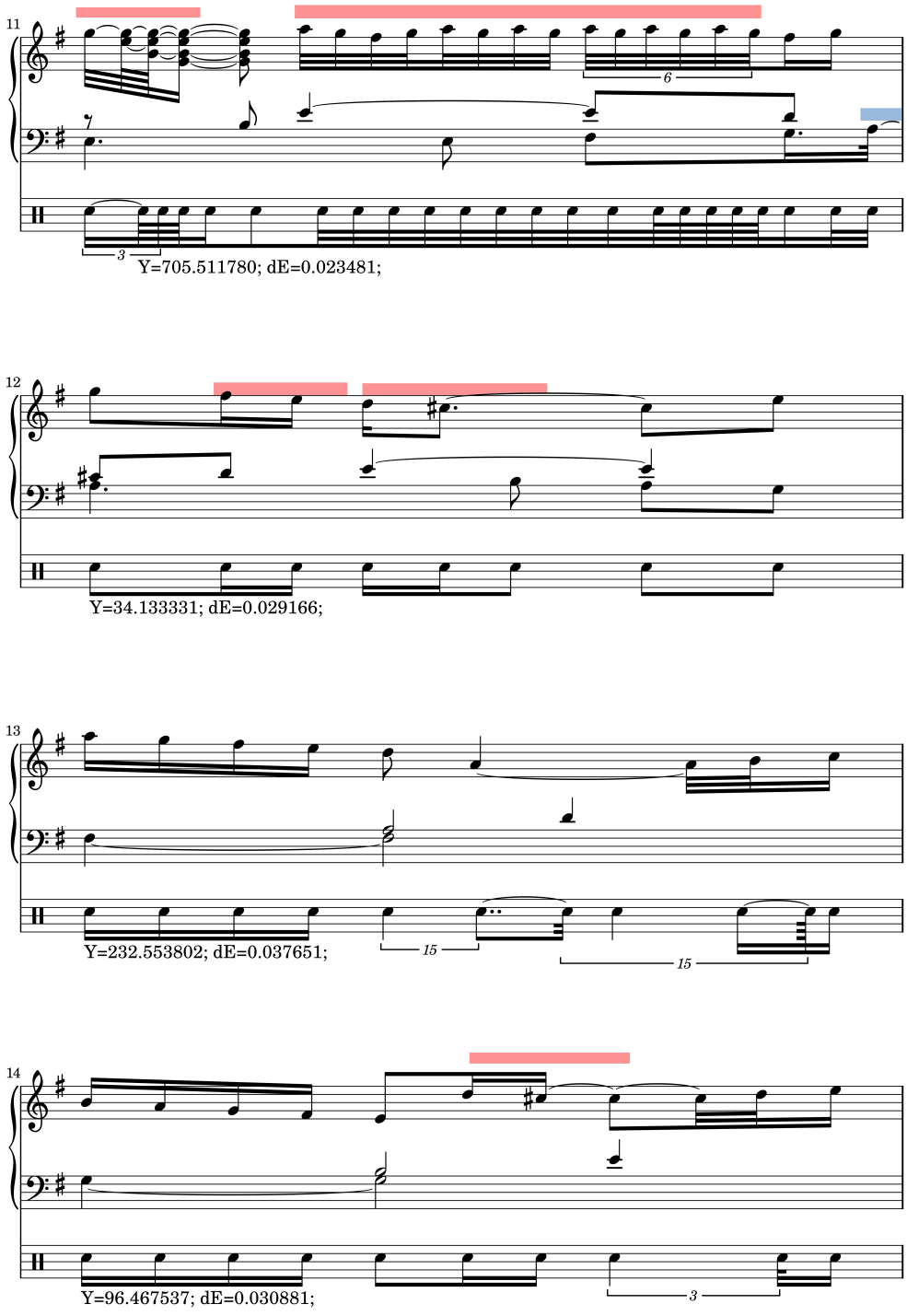


Figure 6-7: Quantisation of Gould's recording from 1981.

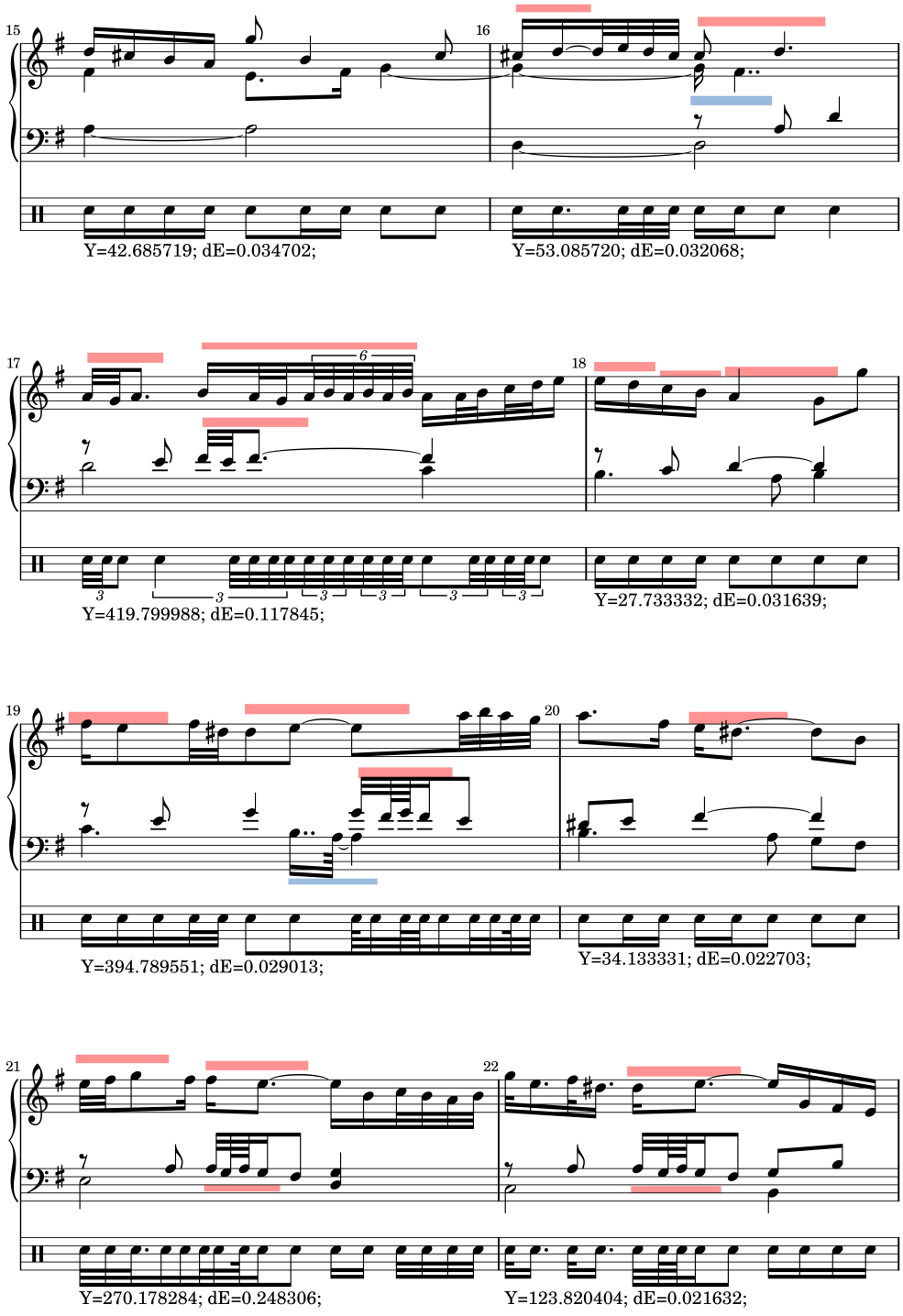


Figure 6-8: Quantisation of Gould's recording from 1981.

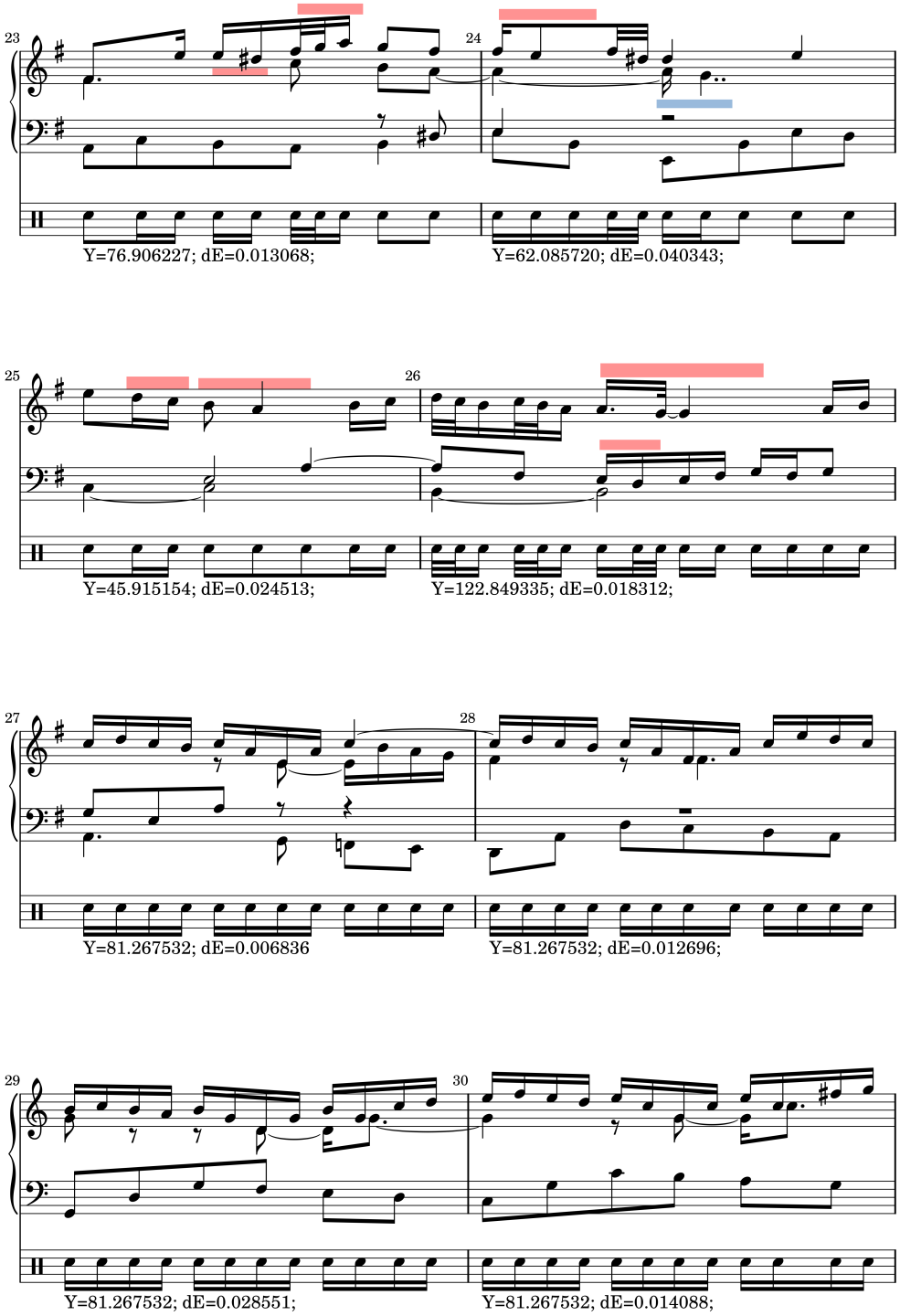


Figure 6-9: Quantisation of Gould's recording from 1981.

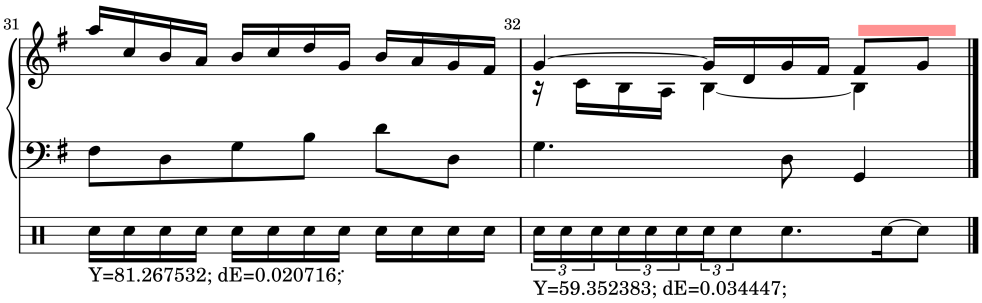


Figure 6-10: Quantisation of Gould's recording from 1981.

Chapter 7

Retentional Maps of Rhythms and their Use for Composition and Music Analysis

This section is an expanded version of our recent ICMC paper, see Boenn (2008). We are reporting the generation of rhythmic structures from single lines to polyphonic voices based on the ideas of Edmund Husserl's phenomenology of inner time-consciousness, see Husserl (1966), and how this relates to our model of rhythm and metre using filtered Farey Sequences.

Starting with Husserl's seminal text, we would like to explore first how musical perception is based on the conscious reduction of acoustic and spatio-temporal diversity into unity. In his lectures Husserl exemplifies his views with an analysis about the perception of a single sound and a sequence of sounds within a melody. They deliver an informative basis for us in order to re-think the perception of musical time and the perception of rhythms and beats in particular. We demonstrate the impact of his phenomenology on music analysis and composition.

Husserl's lectures have been very influential on musicians and composers during the past century. We mention the conducting school of Sergiu Celibidache, especially his student Markand Thakar, see Thakar (1990) and Thakar (1999). Furthermore, we see references to Bergson in Susanne Langer (1953) and we find a discussion of her views on 'time as passage' and references to Koechlin and Bergson in an essay by Elliott Carter who also finds his thoughts related to Husserl, see Carter (1997) and Schmidt (1999). It is also noted that the Berlin school of Gestalt-Psychology has its roots in the school of Brentano, with whom Husserl, Carl Stumpf and Christian von Ehrenfels studied at the same time. There are numerous references to Gestalt-Psychology in the wider musicological field (Deutsch, 1999b; Tenney, 1992; Schenker, 1969).

7.1 Retentions and Protentions

Husserl's term of retentional consciousness (Husserl, 1966) is related to what cognitive psychology later described as working memory. The counterpart to the retentions are the protentions that point towards the horizon of expectations. The origin of the protentions are the now-points. Protentions are expectations about the immediate future. They are born immediately out of the now-point and are derived from the current perceptions, which reside in the impressional consciousness. The consciousness compares permanently the contents of the perception originating from the now-point with retentional phases of earlier content within the working memory. Those phases sink further down into unconsciousness the more distance is created between the temporal position of a phase and the now-point. On the other hand, there is a comparison process between the current perceptual content and past protentions. We assume that learning processes originate from here, which influence the generation of further protentions.

Recently, David Huron (Huron, 2006) drew from the current body of research in cognitive musicology with regard to learning and expectation processes in music listening in order to develop his own ITPRA theory[1]. Huron's focus is on the emotional responses to expectations that root deeply in the history of human evolution.

Based upon Husserl's Phenomenology and upon notions of the Gestalt-Psychology we programmed tools that analyse music recordings in search for temporal expressive profiles of performances and which generate transcriptions representing the underlying score, see chapters 5 and 6. To achieve these goals we had to develop a deeper understanding about the perception of rhythmic Gestalten and how musicians express themselves through timing. This leads us now also to notions of self-organisation, which feed back into composition techniques of polyphony and variation. This shall be explained with examples in the following sections.

7.2 Onset Rhythms

The onset of a sound is very important for the perception and recognition of musical timbres and the learning of timbre categories. "[...] Researchers [...] have found that [sound source] identification relies on onsets. Identification is accurate if listeners hear onsets, and poor if they don't. [...] the lack of onsets made it harder to identify the blown instruments in particular."(Iverson and Krumhansl, 1993). If the presence or non-presence of the onsets are of such an importance it might well be that their occurrences leave markers in the consciousness which enable and support certain types of rhythm processing in the brain.

From the phenomenological point of view an onset is a primal impression from where a field of running-off continua starts to develop that belongs to the entire time-object 'sound'. Husserl states that the now-point is a "creative now", from which a chain of ever new now-points starts. Husserl is then interested in the effects on the intentional consciousness caused

[1]ITPRA stands for the sequence of the following expectation-related responses: Imagination, Tension, Prediction, Reaction and Appraisal.

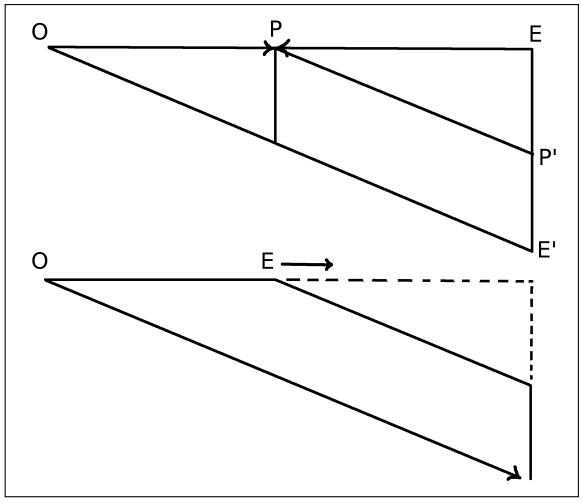


Figure 7-1: OE series of now-points; OE' Sinking-down; EE' Continuum of phases (now point with horizon of the past); E-> Series of now-points which will be filled with other objects.

by the primal impressions. The consciousness creates for every new now-point of the sound a vertical phase continuum of all past now-points of the time-object, it samples at that particular moment the retentional phases within the running-off continua (see figure 7-1 for Husserl's original diagram). The entire process of the consciousness has a *double intentionality*. It means that the consciousness directs itself towards the perception of the sound in its unity as a time-object, but at the same time it is capable of perceiving all minute changes of the sound through the internal time consciousness.

To perceive a rhythm means to relate one sound with other sounds by means of duration ratios and thereby to learn and to recognise explicit rhythm categories (Honing, 2002; Papadelis and Papanikolaou, 2004). These processes are based on the selective division of the time-flow by the attentive consciousness. On the other hand there is duration inherent to the sound, a continuous flow and metamorphosis from one state into the other, a multitude of dissolutions, cross-fades and indiscernible spectromorphological changes. The latter refers to the experience of subjective duration that Charles Koechlin recognised when he reflected Bergson's views (Carter, 1997). Susanne K. Langer said: "The primary illusion of music is the sonorous image of passage, abstracted from all actuality to become free and plastic and entirely perceptible." (Langer, 1953) It is that notion of passage that has been very influential on Elliott Carter. The free development of passage becomes also the prevalent theme within acousmatic music that wants to free itself from all academic and worn-out ways of dealing with sounds and composition (Smalley, 1997).

7.3 Retentional Rhythms

We would like to focus not only on one side of the double intentionality of the consciousness, which is becoming aware of the sounds as streaming entities of passage, but also to consider

the other intention understood by Husserl and also by Hegel (Hegel, 1975), which leads to the awareness of rhythmic relationships between and within sounds, their dynamism and their self-induced organisation.

Husserl speaks of the consciousness of succession. It means that the relationship of two sounds (A - B) sinks down in the consciousness and its content changes as well as its perception through the flow of consciousness, because the consciousness of the individual sounds A and B as well as their relationship remain within the retentions that Husserl calls the primary memory. These are the reference points for all future sounds, whose new relationships are integrated with the past sounds and relations through the retentional memory. Here all sounds are being continuously transformed "the way that" they appear.

The physical inter-onset-time between the sounds A and B is of main importance for the perception of the duration of sound A. Although a pause might happen after A for reasons of articulation – A might have decayed before the attack of sound B – their relationship remains characterised by the full time span between them. Of course, one can connect this pause with both sounds and obtain a relation of three elements. With regard to articulation and phrasing this type of musical silence between notes becomes more stringent and convincing if it is carried out with proportions in mind, proportions that take care of the conscious participation of all three elements in that sequence: A - silence - B. It is also clear that the now-points of the onsets and the perceived durations and relationships between different sounds remain at the same position in time within the retentional consciousness. Only in this way it is possible to learn and to recognise rhythmic categories and their Gestalten. According to Husserl, this is due to the *a priori* condition of the homogeneity of absolute time and has obvious implications for the perception of time objects.

It is now possible to draw a map of the original time field within the retentional consciousness. From the continuum of the now-points we select only the note onsets and inter-onset times. It is further assumed that the memory keeps the time-object 'note', which consists of its onset, its sound continuum and its duration, within a continuous feedback loop. The following example in figure 7-2 illustrates the time field by using a ternary rhythm as its base structure of the original rhythm. Its composed ritardando via increasing note durations is replicated through the entire cycle of all levels together. On the other hand, a composed accelerando would certainly lead to compressed versions of the rhythm on the retentional levels. The density of onsets is increasing at the beginning and decreasing when the perceived rhythm has ended. Short durations within the perceived rhythms are running faster through all levels than long durations. The change of short and long durations leads also to crossovers of the past now-points, i.e., the trajectories of the running-off modes (Husserl) diverge or converge at points that are determined by the durations of the perceived rhythm. At certain moments there are simultaneous onsets, which means the feedback-loops become synchronised. How many synchronisation points happen is again determined by the perceived structure. We think that those synchronisation points between the retentional layers and the perceived layer have an influence on the perception of beat or tactus and that deviations from the synchrony enable listeners to follow tempo changes with adequate response. On the flip side one can say that cer-

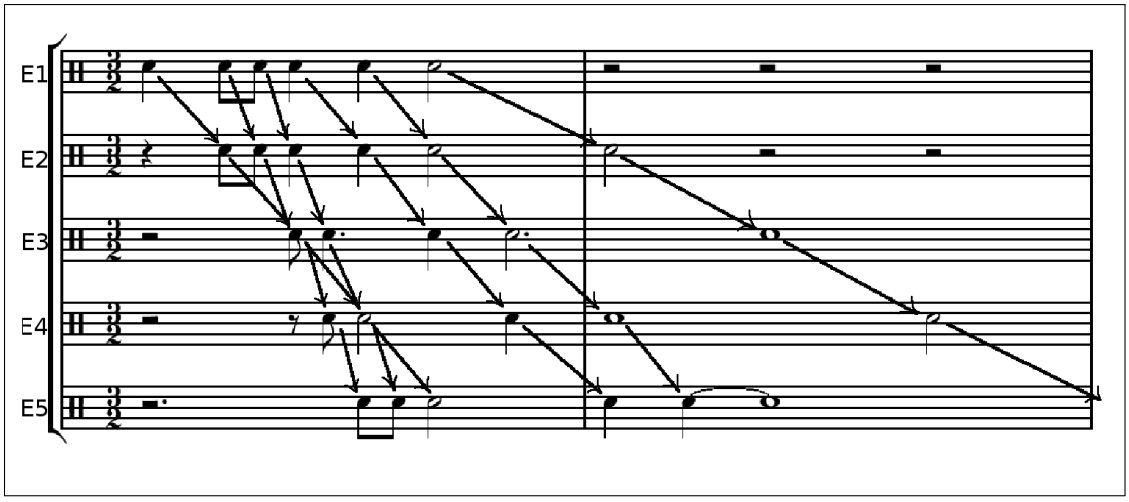


Figure 7-2: A perceived rhythm E1 and its four retentional layers E2–E5

tain rhythmic structures are designed to create a high number of these synchronisation points whereas other designs want to more or less avoid them. At least one can say that because of this structure of internal time consciousness every rhythm perceived, even if it is only a single voice, has an internal tendency to create a polyphonic network of its own. More than that, none of the retentional layer show a simple imitation, a delayed version of the perceived rhythm, but presents us always with a variation. It would be therefore most interesting to study classical polyphonic pieces to determine whether they show any relations between the rhythms of voices that could have been derived from a retentional version of a main voice.

A second example illustrates the properties of the retentional rhythms. The original rhythm in figure 7-3 is taken form Elliott Carter's *String Quartett No. 2* and it is a composed accelerando. The continuous shortening of the durations leads to a focal point, a contraction of the rhythms on the retentional layers. The shortest layer is .75 crotchets long with 4 onsets, whereas the original has 7 onsets within 4.5 crotchets. Beyond this point the structures expand again. We presume that the compression of the first layers reflects and amplifies the perceptual effect of the accelerando.

When investigating how many onsets are synchronised in our example and writing them into a score where the notes on the first staff represent one onset per note, the second staff represents two onsets per note, the third one three onsets and so on, we obtain the following result (see figure 7-4).

We assume that a high number of synchronised onsets within the retentional layers contribute to the perception of a beat or tactus.

It might seem problematic that a notated sequence of durations is only a representation, which is not to be confused with the experience of a real musical performance that always produces unique spatio-temporal phenomena and is therefore unrepeatable. On the other hand, notation is the basis for many Western compositions and performances. Research has shown that rhythmic categories notated in the score are also correlated with a corresponding class

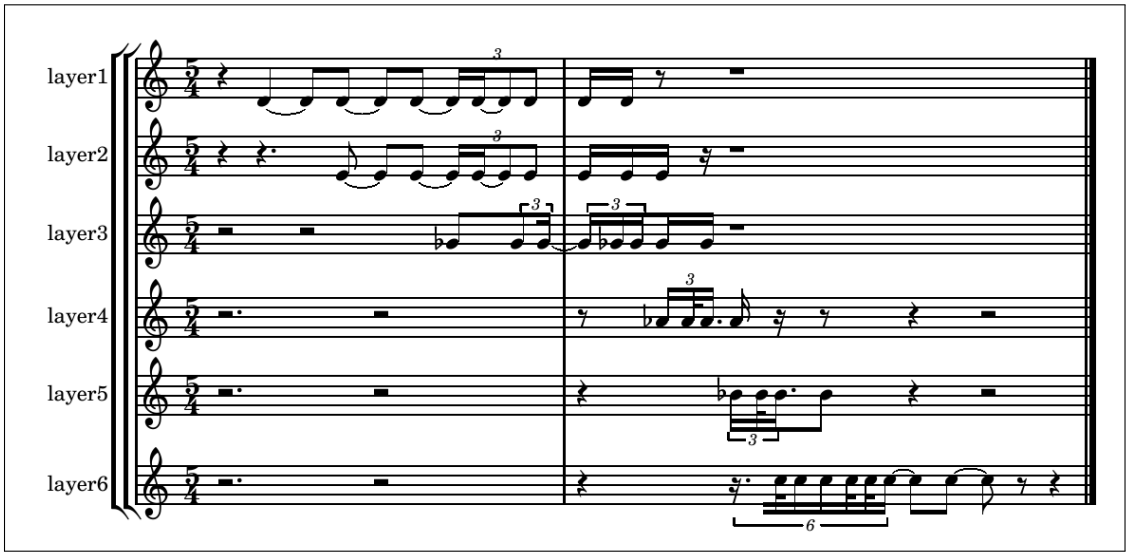


Figure 7-3: The original Carter rhythm (layer1) and its retentional layers.

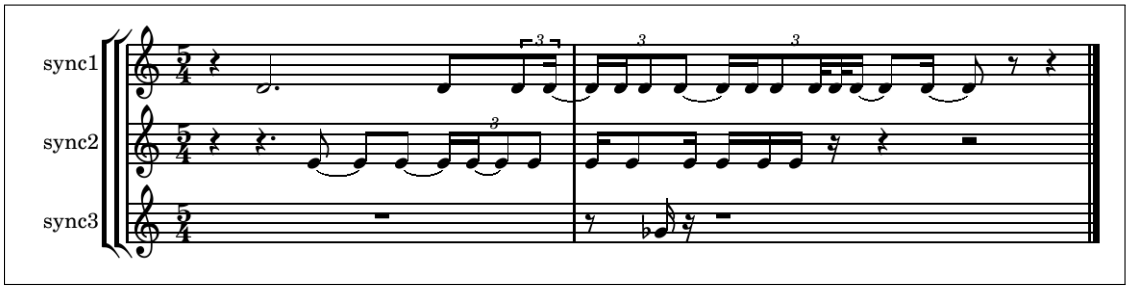


Figure 7-4: How many onsets in figure 7-3 occur at the same time together? The first staff shows onset times that occur only one at a time in any of the retentional layers, the second staff shows onset times that occur in two voices at the same time, the third staff shows one onset time that occurs is three voices simultaneously.

of perceived rhythms (Honing, 2002; Papadelis and Papanikolaou, 2004). It is well known, that even the simplest rhythmic ratio is never played with mathematical precision. But within the cognitive domain they remain simple ratios, because as we have seen every proportion is embedded in a network of relations. And only on the basis of comparisons within the retentional consciousness a listener is able to learn, form and recognise rhythmic categories out of groups and classes of similar proportions. Each of the rhythmic categories represents then a field of possible realisations.

7.4 Compositional Applications

We generate the retentional onset maps discussed so far by starting with a stream of note onsets:

$$S = \{s_1, s_2, s_3, ...s_{|S|}\},$$

Apart from a sequence of notes it is also possible to perceive changing spectral properties as rhythms within a single sound. According to our phenomenological analysis each duration will be kept for a certain amount of time in a feedback loop within the working memory. We present the following examples showing a creative application of our previous analysis. All examples in figures 7-2, 7-3, 7-8, 7-10, 7-12, 7-14 and 7-15, have been generated by using the following equation for the kth layer T_k of the retentional onset map:

$$T_k = \{t_i = s_i + k \times (s_{i+1} - s_i), i = 1, 2, 3, ..., |S| - 1\} \tag{7.1}$$

with $k \in \mathbb{N}$, t_i = onset of the looped note duration within the kth retentional layer, s_i = onset time within the perceived layer S, and $(s_{i+1} - s_i)$ generating the IOIs from the set S.

With k > 0, equation 7.1 generates the onsets of the kth retentional layer. The chances are then relatively high that any t_i may coincide with a $s_{(i+j)}$, with j as an integer $> i$. But if k is a rational number > 0, the chances are relatively high that any t_i may not coincide with a $s_{(i+j)}$, but rather falls in between inter-onset times of the perceived layer. Instead of discrete retentional layers (with k as an integer) there is then a time field with an infinite number of layers between the discrete layers. If k changes for for each T_k one obtains proportional transformations of the inter-onset times within a retentional layer. If k changes for every t_i within each of the individual layers T_k, one obtains non-linear trajectories between the onsets of the feedback loops leading downwards from layer to layer. For example, for each t_i in equation 7.1, instead of doing $t_i = s_i + k \times (s_{i+1} - s_i)$ one could add a small random number ρ like this: $t_i = s_i + (k + \rho) \times (s_{i+1} - s_i)$. This non-linearity might be desirable for a compositional application of retentional onset maps.

If $k < 0$ one obtains the inverse of the principle, i.e., a given rhythm would be mapped to a fictitious rhythm in the past, as if the fictitious rhythm in the past would be the 'perceived' rhythm and the given rhythm S would be a retentional layer T_k of that fictitious rhythm. This negative principle again offers the same options as before, i.e., k as a negative integer or negative rational number, changing from one onset to the next onset or between individual layers. In

the end one can imagine free trajectories through any point of the retentional time field. In this manner one generates new rhythms over and over again, but always based upon one and the same single line of durations.

The retentional onset map $\{\Omega\}$ consists of S and all retentional layers T_k. If one substitutes S with T_0, because equation 7.1 returns S when $n = 0$, one can write:

$$\{\Omega\} = \{T_k, k = 0, 1, 2, ..., \kappa\} \tag{7.2}$$

with κ representing an unknown perceptual limit. In theory, $\kappa = \infty$, but then the divergence of the running-off continua would generate very large durations, which might not be practical from a creative musical point-of-view[2]. In terms of musical perception, a threshold of approximately 6 seconds has been indicated as the limit of the psychological present (James, 1950; Pöppel, 1972; Michon, 1978; Fraisse, 1984), which according to London (2004) would coincide with the duration of a bar in ternary metre with three slowest beats at 30 BPM, see section 2.4. The refinement of equation 7.1 in such a way that the outcomes of T_k are bound to the perceptual limit of psychological present is left for future work.

7.5 Retentional Rhythms and Farey Sequences

The retentional onset map $\{\Omega\}$, see equation 7.2, which consists of all sets T_k generated after equation 7.1, is a filtered Farey Sequence. If the set of onsets S is normalised and the output of a quantised musical performance, then the set S itself would be a filtered Farey Sequence. Even if S is an unquantised normalised set of onsets, it can still be regarded as a filtered Farey Sequence, although it would be of much higher order, depending on the highest denominator amongst the rational numbers in S that are represented by their continued fraction expansions. Therefore, S can always be regarded as some filtered Farey Sequence F_n, with n representing the order of the Farey Sequence.

We give here an example of how one can analyse $\{\Omega\}$ in terms of the rational numbers, which are contained in a filtered Farey Sequence. Figures 7-5 and 7-6 show how individual note onsets coincide with fractions contained in F_{12}. The examples are based on a retentional map of rhythms shown in figure 7-2. The combined set of fractions generated with this procedure is a filtered Farey Sequence F_{12}. The mapping is produced by comparing the normalised onset times on all voices per bar with F_{12}.

One can conclude that the algorithms described with equations 7.2 and 7.1 generate retentional onset maps, which always coincide with fractions of filtered Farey Sequences.

[2]It might be interesting for renderings of John Cage's piece *As Slow As Possible*. One performance of the piece has been designed for a church organ in Halberstadt, Germany. The concert started in February 2003 and will continue for another 631 years, see `http://www.john-cage.halberstadt.de` [Accessed September 2010]

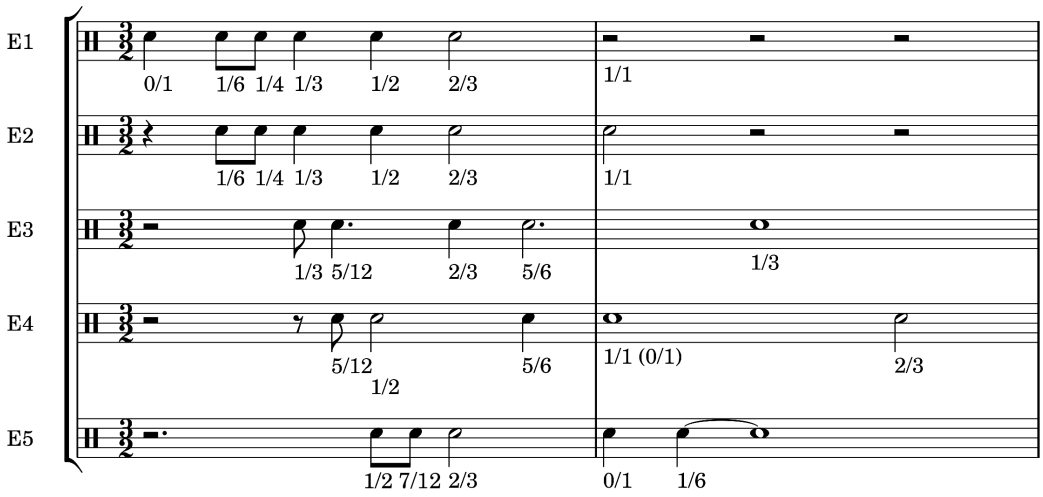


Figure 7-5: Onsets in retentional maps coincide with fractions of a filtered Farey Sequence F_{12}.

7.6 Examples

Figure 7-7[3] shows another example of synchronised retentional rhythms using the first eight bars of the *Aria* from Bach's *Goldberg Variations*. When evaluating the points of synchronisation within the retentional time-field it becomes evident that the incidents of synchronisation on the lower levels are often identical with the beat onsets of the 3/4 metre. This leads us to the assumption that the synchronicity of onsets within the retentional layers together with the original perceived rhythm delivers important cues to the brain so that it can perceive the beat-structure (or tactus) of a musical performance. But we are not yet sure whether this assumption will hold true on the basis of onsets extracted from real performances. It remains one of the future tasks to investigate that possibility on the basis of real-world performance data.

The following example in figure 7-8 is based upon Debussy's piano piece *Gollywog's Cakewalk* from his suite *Children's Corner*. It shows us an interesting hoketus-like retentional rhythmic structure, which originated from the syncopations of the ragtime rhythm. The synchronisation score shows also a manifold of various attack densities, see figure 7-9.

It is interesting to see in Ravel's *Bolero* that the last quaver of bar 1 is amplified by the first four retentional layers to become an especially energised upbeat-impulse towards the second bar, see figures 7-10 and 7-11.

In a situation where the original rhythm is composed as an ostinato, it will sooner or later re-appear on one lower retentional layer, see figure 7-12 using a 7/8 metre.

In contrast with the above examples, there are almost no synchronisation points in a rhythm of randomly generated durations, see figure 7-13.

[3]The pitch information shown here and in the following examples is only included to facilitate reading, the pitches have no other significance and are not related to the original score.

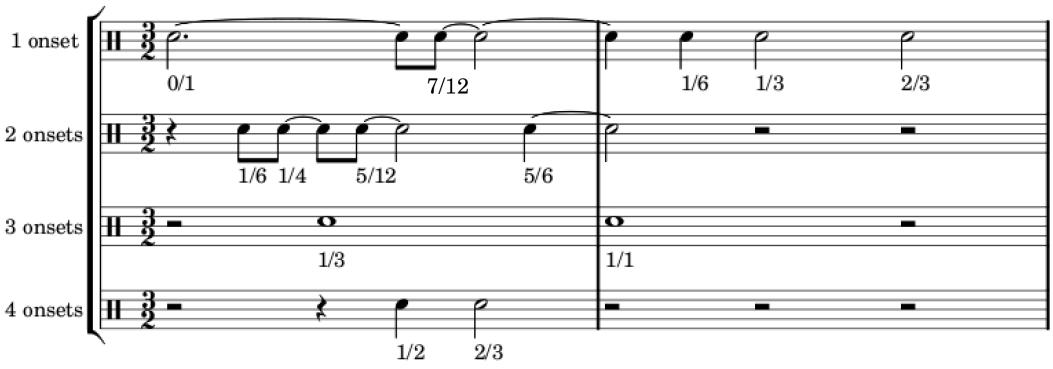


Figure 7-6: Here we show the synchronisation map of example 7-2 where its onset points coincide with fractions of a filtered Farey Sequence F_{12}.

As an example of using equation 7.1 we show in figure 7-14 a projection of the Carter-Rhythm shown in figure 7-3 using $n < 0$, i.e., the Carter-Rhythm on layer 1 is the last retentional layer of the rhythm in layer 13, which started in the past, see figures 7-14 and 7-15.

7.7 Retentional Rhythms and Neuroscience

Recent neurophysiological research suggests that the brain uses the very same neural structures for counting *and* timing (Meck, 2003). The unified mode-control model states that the abilities of timing and counting are based on a pacemaker-accumulator system. Within such a system, numbers of events as well as their measured time-spans are represented by magnitudes in the accumulator part of the system. One has therefore reason to speculate that if an internal quantisation towards simple ratios exists it would be greatly supported by this neural structure. For our model of retentional rhythms it would mean that event onsets can be counted, for example a stream of pulses induced by the brain from oscillations of the perceived event durations, i.e., the retentional onset map. According to the mode-control model, the time elapsed between perceived and induced event onsets could be measured using the same neural structures. We speculate that there is a feedback loop that induces the best fit of an underlying pulsation grid to the measured pattern of perceived event onsets. It seems also likely that the magnitudes of perceived individual durations could easily cluster together in order to form patterns of simple duration ratios like the ones we meet in Western and non-Western musical cultures. Being confronted with a manifold of different magnitude ratios on many different time-scales, within the boundaries of human perception, it turns out that the simplification of this manifold towards simple integer ratios also guarantees that the pattern extracted from this manifold can be stored in memory very efficiently. Instead of storing millions of rhythmic patterns which all come close but are not equal to a set of simple ratios, it makes more sense to store only the simple ratio pattern and to have a process running that is based on Gestalt recognition principles and that works on the rationalisation and quantisation of the incoming data. The

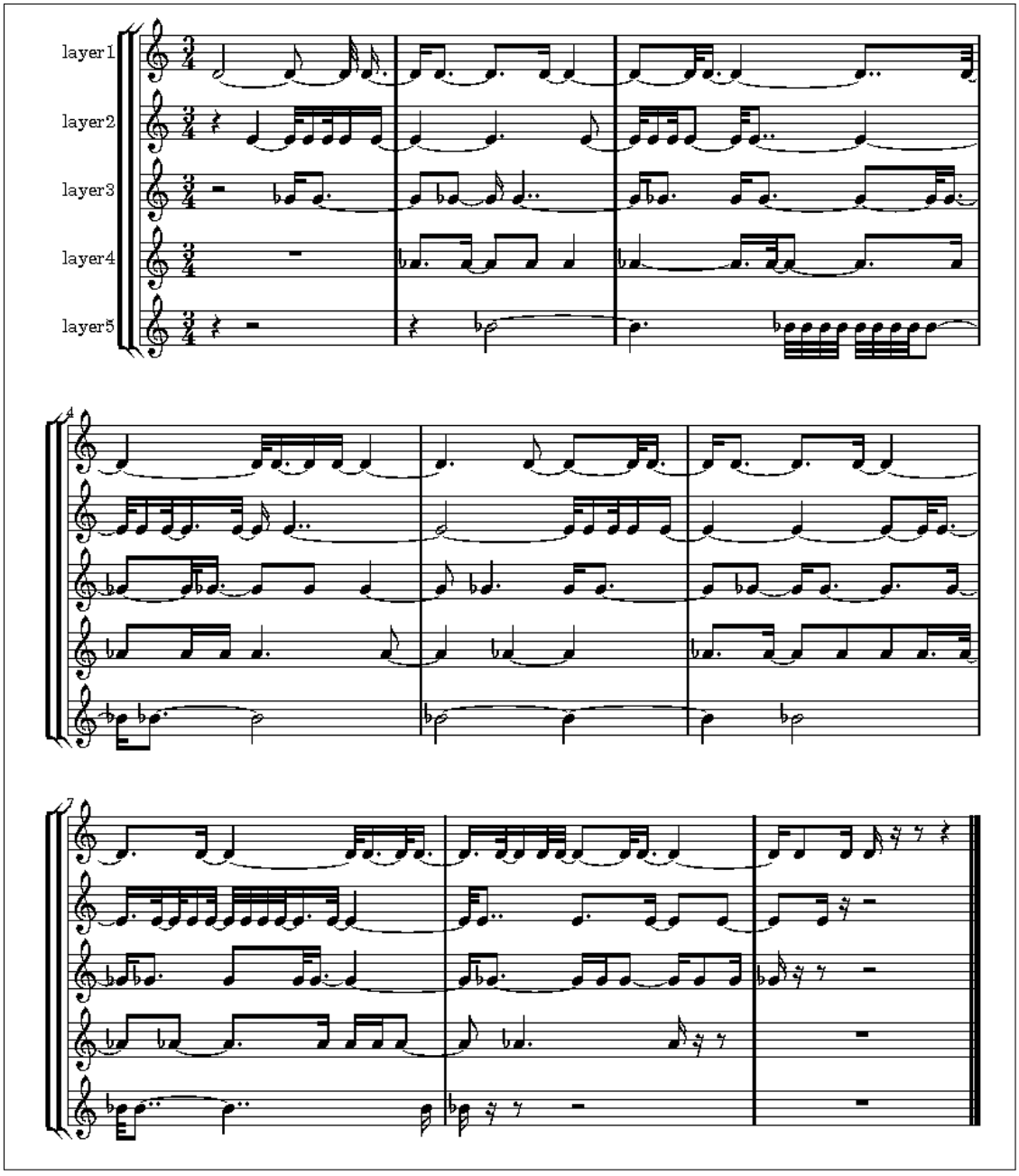


Figure 7-7: The first eight bars of the *Aria* of Bach's *Goldberg Variations* with ornaments notated metrically. This is the polyphonic score of the synchronisation points of the retentional rhythms constructed from the original compound rhythm of both hands. Layer#1 represents one attack point, layer#2 represents two attacks, and so on. The material is based on a retentional score with four layers.

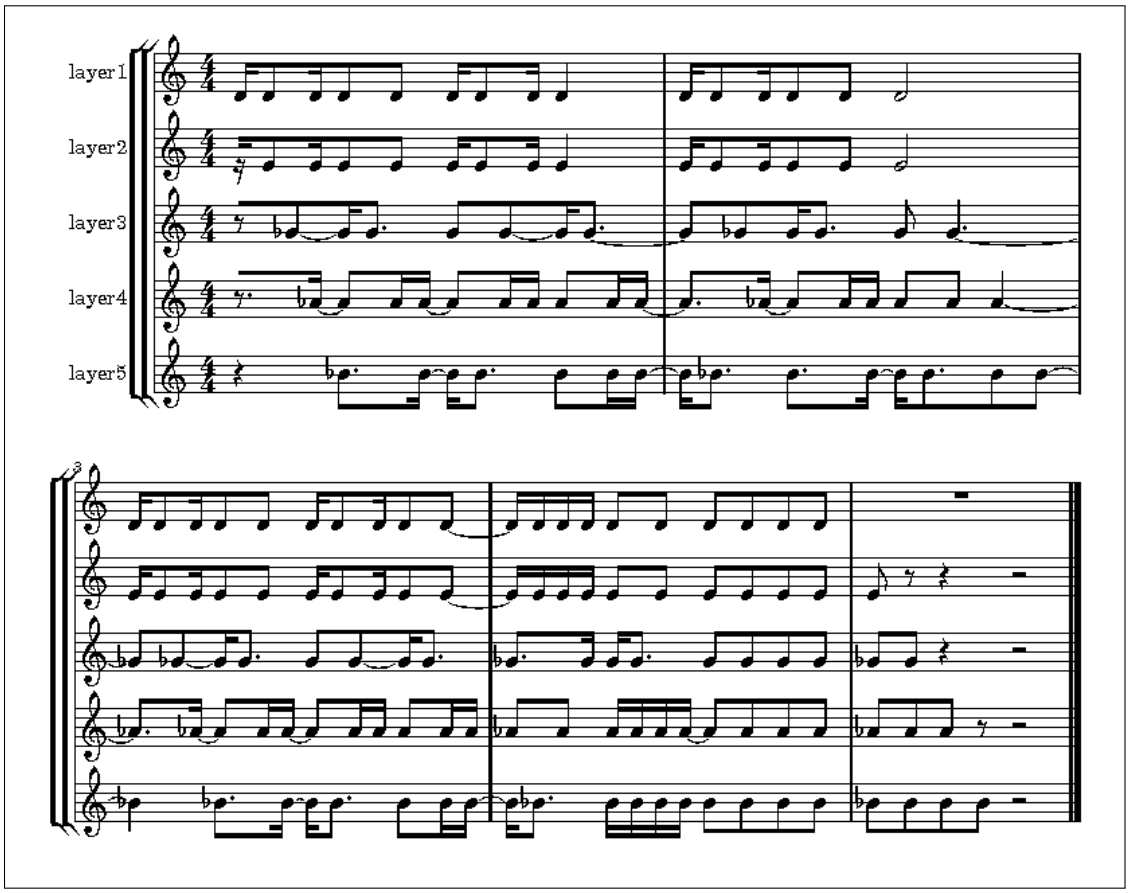


Figure 7-8: The first bars of Debussy's *Cakewalk*-Rhythm in the form of a retention score.

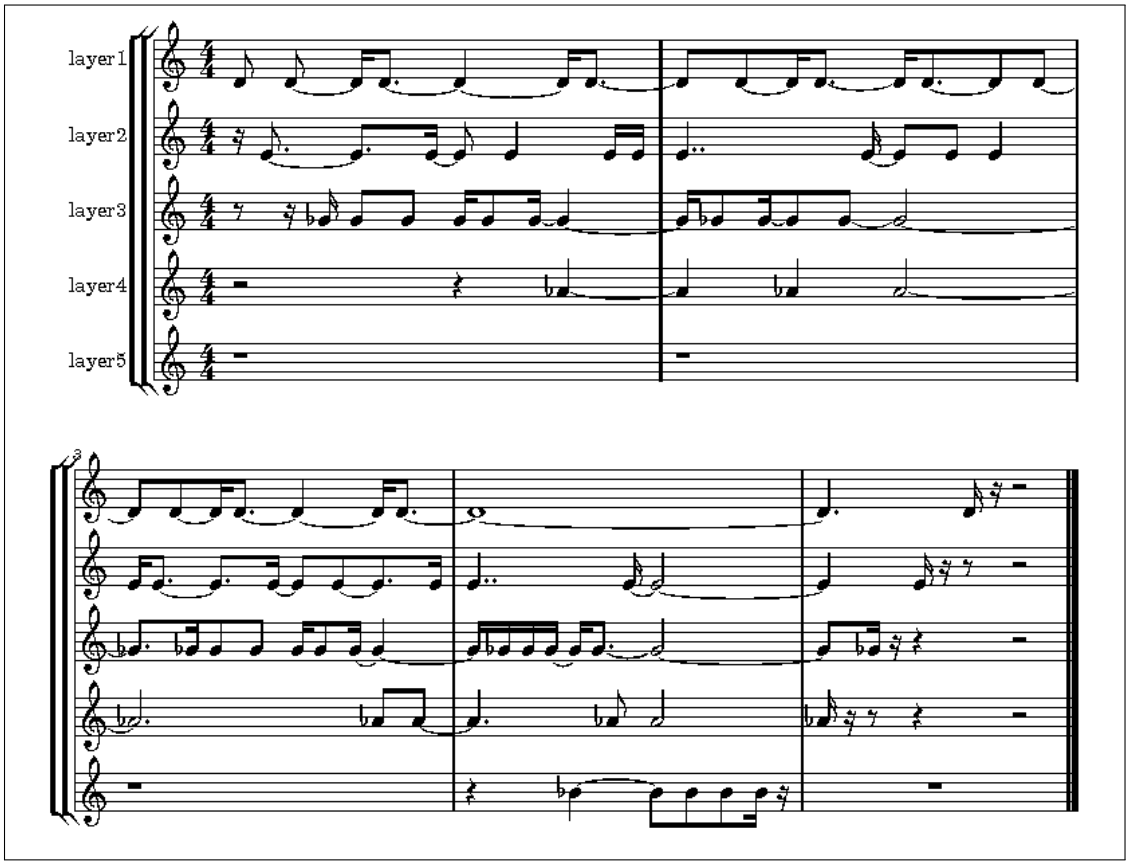


Figure 7-9: The example of Debussy's *Cakewalk* (see figure 7-8) shown as a score of synchronisation points.

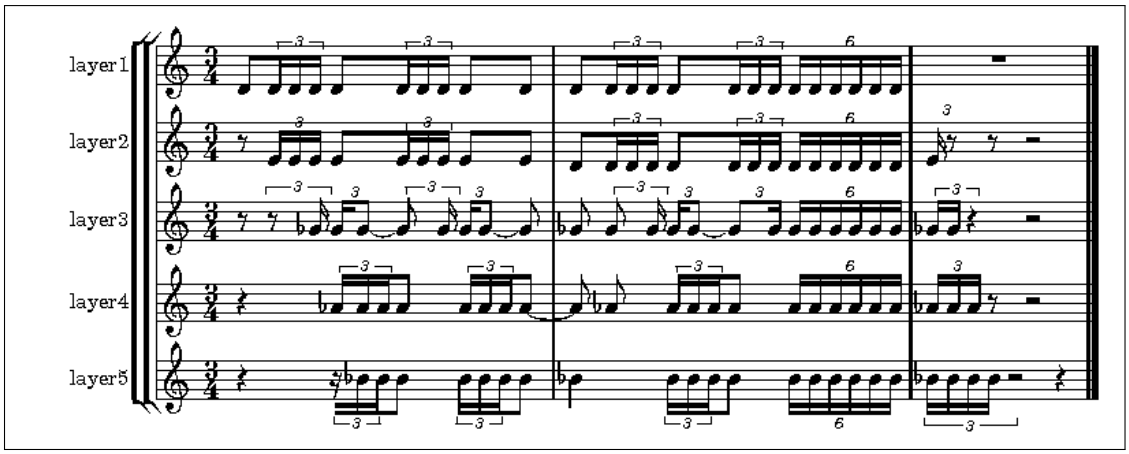


Figure 7-10: The two-bar *Bolero*-Rhythm shows a remarkable synchronicity on the last quaver of bar one and through the last group of semiquaver triplets in bar two. In order to demonstrate the extraordinary composition of this famous rhythm it is enlightening to precisely swap the sequence of the two bars and to add the melody again. It will suddenly make the different effect very clear that those impulses have on the last beats. The first bar gathers dynamic energy whereas the second bar releases the energy again.

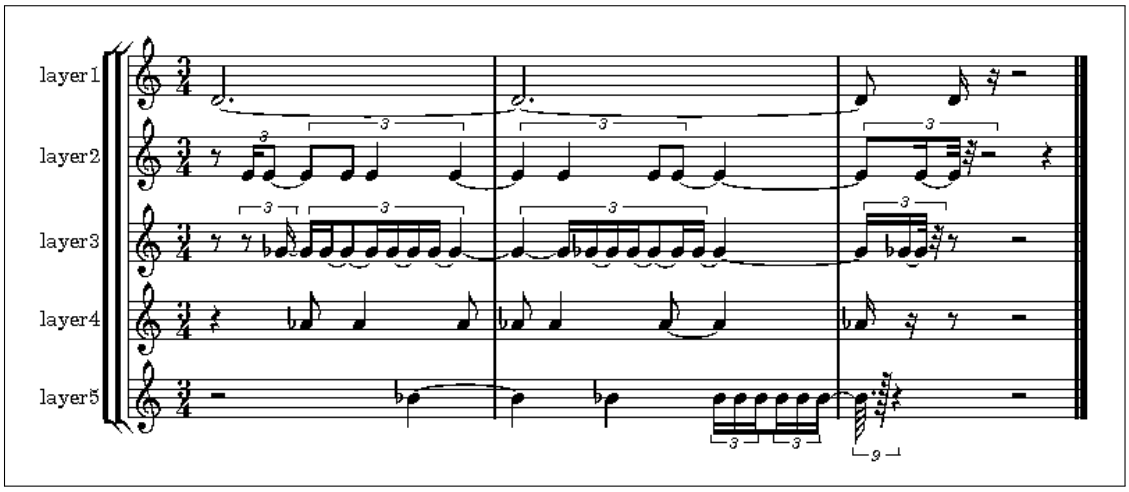


Figure 7-11: The slower pulsation of the lower layers, which correlates with the orchestral accompaniment of the rhythm, becomes evident in the synchronisation points between the retentional layers of the *Bolero*-Rhythm.

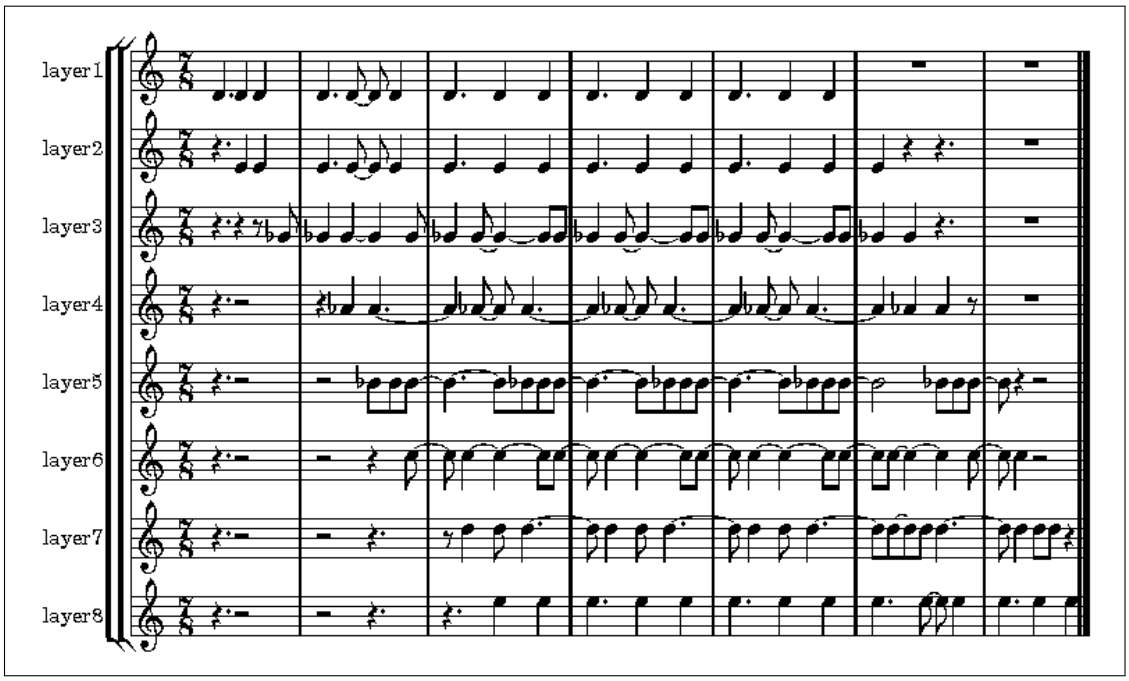


Figure 7-12: The perceived ostinato rhythm in 7/8s re-appears on the 7th retentional layer.

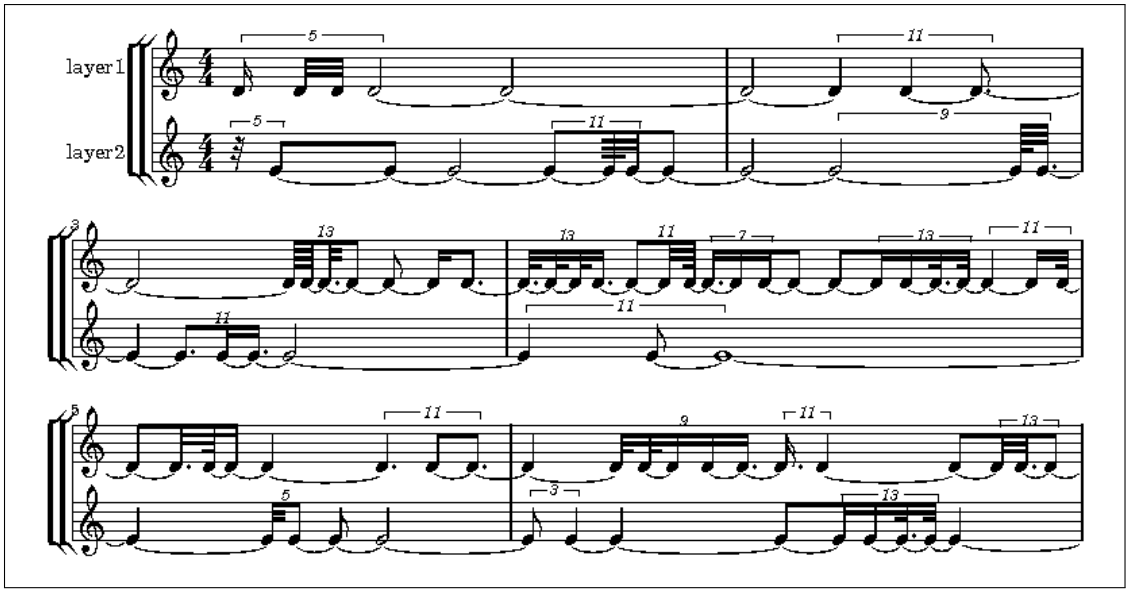


Figure 7-13: As expected in a randomly generated rhythm there is no correlation with periodic isochronous beats. Here is the synchronisation score with the first voice representing one attack, the second voice representing two attacks. The retentional score had five layers, original rhythm and 4 retentional layers. The frequency of the double attacks represented by the second voice originates from the definition of duration that is equal to the inter-onset-interval. Therefore, the first retentional layer is always synchronised with the original rhythm.

bespoke neurological system makes it also simple to reproduce the recognised patterns, because timing and counting are based on the same method in the brain and thus simple integer ratios become easier to reproduce than choosing and reproducing from a manifold of numerically more complicated patterns. This is not to say that more complex patterns are totally excluded, they are just better integrated and manageable in a rationalised way, i.e., because of the ability to form rhythmic categories more complex patterns are integrated into a formal hierarchy together with simpler ratios, very similar to the structure that unfolds in the Farey Sequence, which we consider as the grid structure for quantisation but also for generative rhythmic processes in performance and composition.

7.8 Conclusion

Retentional onsets maps are an original way of applying Husserl's philosophical notions about the perception of musical time in the domain of music composition, where the development of rhythms and musical form are an important area of a composer's creative work, see section 1.1. The idea that musical structures are linked together in an intricate, generative way, and further that a present content of perception originates from the experience of all musical events, which happened before, up to the present moment, is strongly supported by Husserl's phenomenology. Retentional onset maps are an interesting tool for the exploration of how musical rhythms are

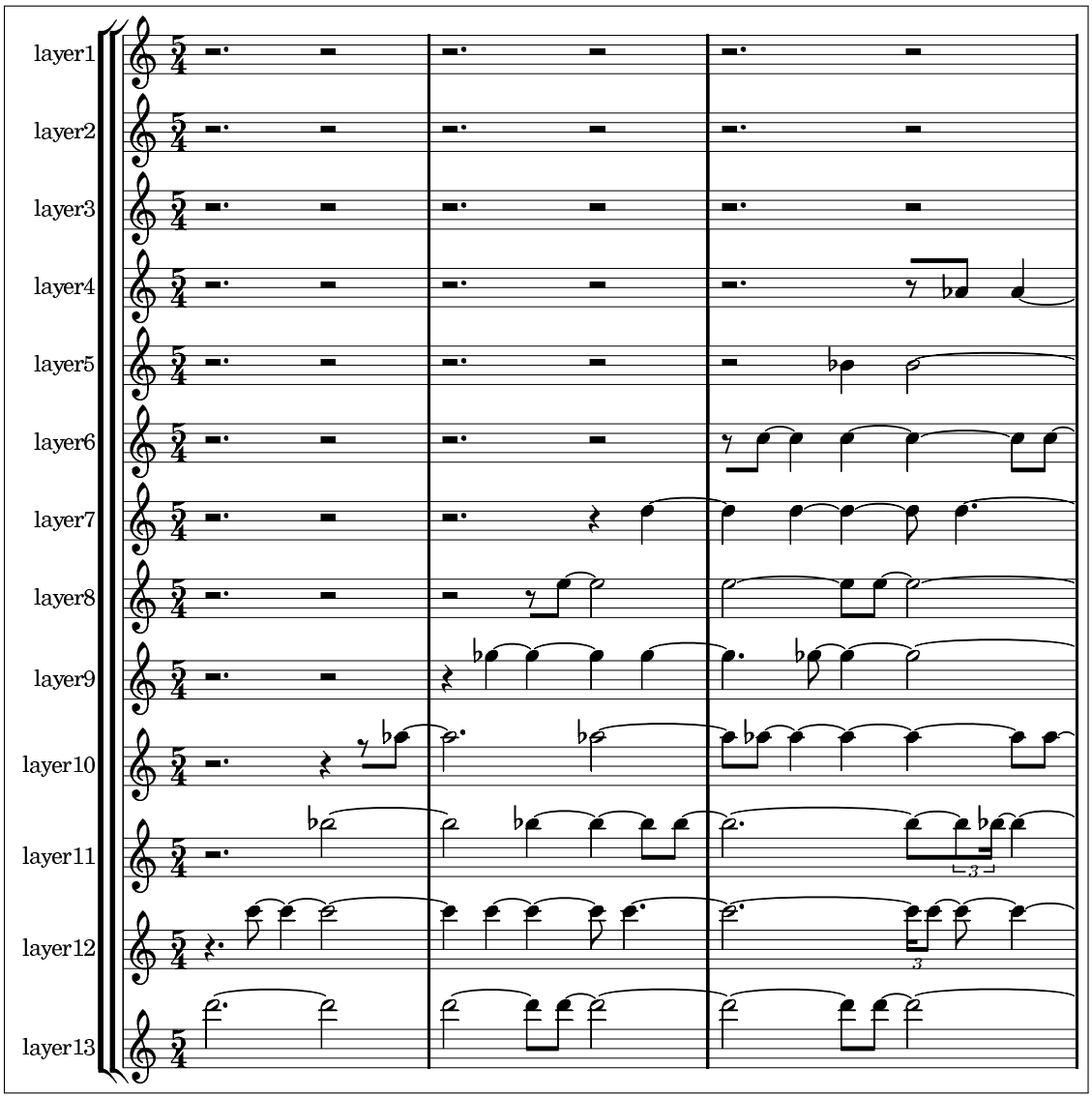


Figure 7-14: Example of equation 7.1 used on the basis of the Carter-Rhythm, see figure 7-3 with $n < 0$. The "original" rhythm is on layer 13. The process ends in layer 1.

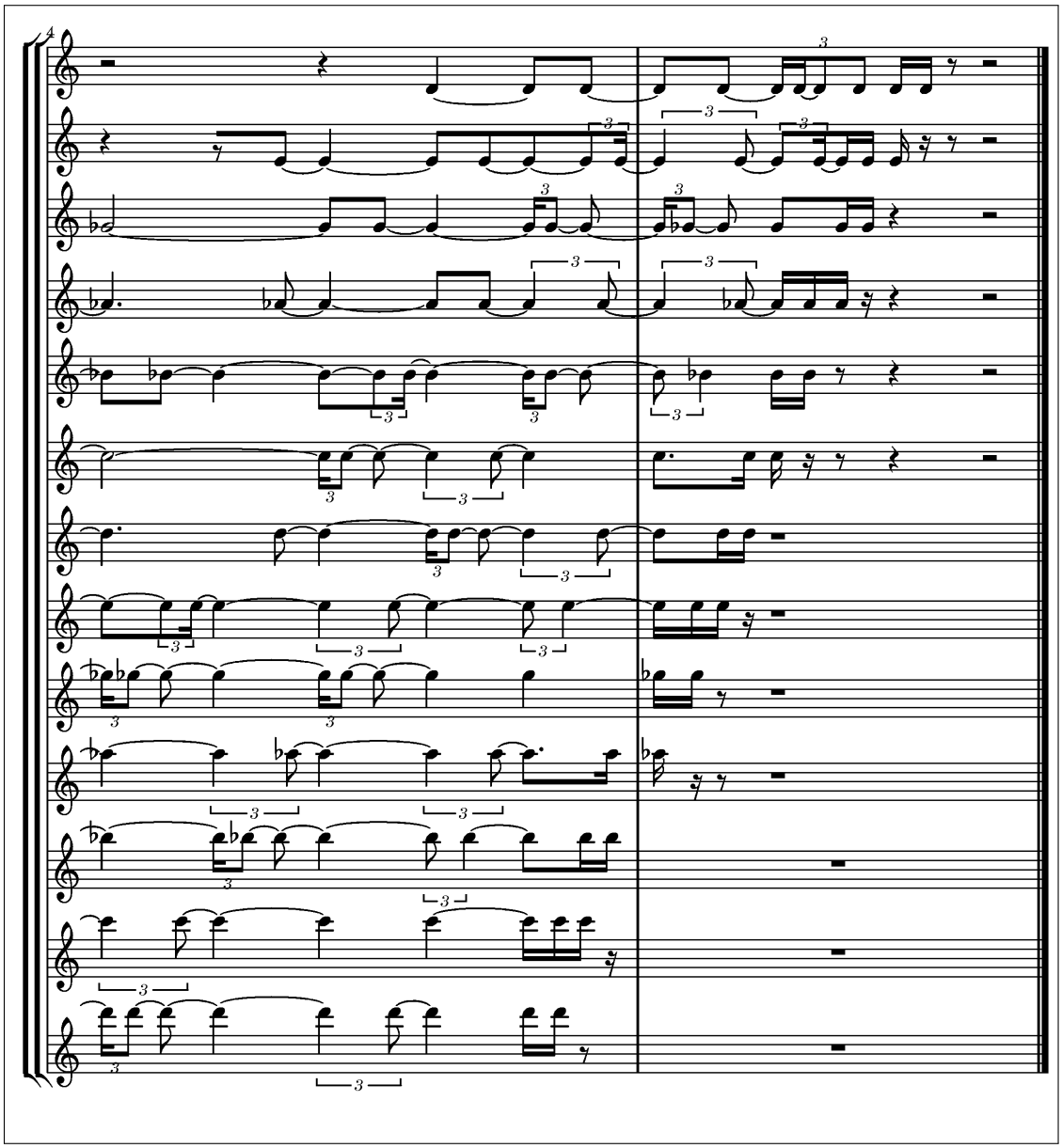


Figure 7-15: Figure 7-14 continued.

kept in short-term memory.

In the philosophical domain Husserl gave us a lot of insights into the consciousness of a listener when perceiving sounds and sequences of rhythms. We have shown the impact that Husserl already had on many musicians in the past century and pointed to important questions that still remain, i.e., could we prove in future field experiments the existence of retentional rhythms? We also demonstrated a huge potential for musicological research on rhythm composition and discovered also a new organic way of dealing with rhythms that would be useful for composers and musicians.

Chapter 8

Future Work

8.1 Introduction

In this chapter we would like to point to a few lines of research that can evolve from the results that we have presented in our thesis. First we would like to reflect on the retentional onset maps discussed in the previous chapter and how they could be applied to musical performances and to the analysis of the onsets that can be extracted from them. Secondly, we would like to describe how we could integrate the components for window segmentation, grouping, quantisation and transcription into a complete system.

8.2 Retentional Rhythms

We would like to work on the refinement of equations 7.2 and 7.1 in such a way that the outcomes of $\{\Omega\}$ and T_k are bound to the perceptual limit of psychological present. We believe that these equations could then work well as a model to describe the content of the perceived timing informations that reside in our memory during this window of time. In addition we had the idea of extracting information from $\{\Omega\}$ in terms of those onsets that converge to a close onset position on a particular retentional layer T_k. We have made some initial experiments using different sets of onsets S from musical performances, for example we have tested both Gould recordings of the Bach *Aria*. The problem we have faced is related to the inherent tempo fluctuations of musical performances. These tempo modulations will shift the subsequent trajectories onto retentional layers. The very focussed points of synchronisation that we have analysed using CPN scores are getting blurred and it is becoming more difficult to extract the points of synchronisation from the surrounding stream of onsets. Hence it seems to be difficult to track local tempo changes.

Help for this problem might come from our grouping, quantisation and transcription algorithms, which have been successful in detecting local tempo changes due to expressive timing and rubato. If we would group onsets on each retentional layer T_k and possibly also quantise the means of the duration classes detected by the grouping, then it might be easier to find the

points of synchronisation between various retentional layers. In our opinion, this seems to be a viable line of research.

It might find additional support if field experiments for the perception of retentional rhythms could be carried out. Musicians could be confronted with 6 seconds long stimuli containing neutral clicks from a monophonic rhythm. Various tempi according to a metrical tempo grid should be used. After the stimulus stopped, the test group could be asked to tap not the same rhythm but a possible improvised continuation of the stimulus they have just heard. A structural analysis and comparison of the tapping results with the stimuli could reveal salient points from the retentional onset maps of the stimuli.

Because of the structure of the retentional onset maps, every monophonic rhythm has an inherent tendency to create a polyphonic network of retentional voices. None of these retentional layers is a simple imitation, a delayed version of the perceived rhythm, but presents us always with a variation. It would be interesting to study classical polyphonic pieces of the Renaissance, where the constant variation and non-repetition of rhythms is quasi a rule for composition, in order to determine whether they show a rhythmic structure that could be similar to a retentional onset map of a particular voice.

8.3 Quantisation and Transcription

We have presented the building blocks for a working system for the quantisation and transcription of compound rhythms generated by note onsets from polyphonic musical performances. We are now developing an integrated system. The plan for it involves one or both automated window segmentation processes in order to generate analysis windows from a large initial set of onset data, see sections 5.3 and 5.4. The grouping, quantisation and transcription algorithms described in chapters 5 and 6 then generate quantised solutions for each of the windows. We are working now on a stitching algorithm, which would ensure that all of the quantised windows would form a continuous large set of quantised onsets ready for transcription into CPN. The main idea is to take the last quantised duration of an analysis window and to form a repetitive stream of events with that duration, which we project further into the segment of time covered by the next window. We will then count and collect matches of the projected set of onsets with the quantised onset positions inside this window. We can discover with this procedure the tempo relation between the two quantised windows. With this information it would be possible to unifiy the durations of all subsequent analysis windows under a common reference value needed for CPN. If this succeeds then we can stitch all quantised windows together to a continuous score in CPN. We have run some initial tests, which have shown encouraging results.

Chapter 9

Summary and Conclusion

In this thesis we have shown that musical rhythm and metre can be modelled by using filtered Farey Sequences. The Farey Sequence is particularly useful because it can be easily mapped to different perceptual and musical timescales. Unlike CPN, a Farey Sequence is not bound to a common reference duration, such as the semibreve. It is therefore not simply bound to the concept of beat subdivisions and models of musical metre, it can also serve to analyse and to generate large musical structures, as we could see for example with the analysis of Igor Stravinsky's *Procession of the Sage*, see section 3.3, or with the possibility of sequencing Farey Sequences together in order to form a filtered Farey Sequence of higher order, see section 3.2.

We have learned in chapter 2 that Western music notation uses small-integer ratios for the encoding of note durations, rests and musical metres. Metre is a cross-cultural phenomenon of entrainment (London, 2004). Based on the analysis of cyclical patterns from different cultures, for which an extended necklace notation is used, London (2004) arrives at well-formedness constraints (WFCs), which describe the general structure and perception of musical metre. He also poses a *Many Meters Hypothesis* (MMH): Human listeners have memory of various tempo-modifications, rubati, and expressive timings, that are associated with a particular metrical pattern. This association forms via entrainment and through learning within a particular cultural framework.

The successful quantisation and transcription of Bach's *Aria* performed by Glenn Gould provides supporting evidence for the MMH. The quantisation per bar to elements of a Farey Sequence suggests that although the metrical framework is hidden behind a rhythmic musical surface layer, it can be revealed and the expressive timing that is given to the metrical framework can be filtered out and subtracted from the performance data. Listeners will always experience a slightly different rendering of the same musical piece, even under similar conditions. Therefore, by using expressive timing, a performer will always bend the metrical framework by various amounts, every time he or she performs the piece. A listener is then always confronted with variations of the same metrical framework, yet one is still capable of perceiving the very same metre. In order to model this process we have devised a grouping algorithm, a quantiser and a transcription algorithm. The score of the performance can be unknown to the program.

The adjacent interval spectrum by Kjell Gustafson (Toussaint, 2004) proved to be useful in order to recognise Gestalt principles amongst inter-onset intervals. These principles led to the development of a grouping algorithm, which successfully groups IOIs into duration classes, see chapter 5. Using our grouping method and the detection of proximity and similarity between note durations in adjacent interval spectra we are able to detect duration classes within an analysis window. We found out via experimentation that these analysis windows cannot become infinitely large. They have to be bound by a selection of perceptual timing limits.

IOIs have been extracted from audio recordings using an onset detection program library, *aubio*, by Brossier (2006). We have learned that onsets are sufficient for the communication of musical rhythms and metres (Large, 2008). Onsets can be also extracted from MIDI files or real-time MIDI performances. Alternatively, the recording of manual tapping on a computer keyboard is also a viable method for the generation of experimental data sets.

We have put the grouping, quantisation and transcription algorithms to the test with two recordings of the *Aria* of the *Goldberg Variations* by J.S. Bach played by Glenn Gould in 1955 and in 1981. The challenge of the *Aria* lies in its many ornamentations and syncopated rhythms. In addition, we had to deal with polyphony in two hands, which produces compound rhythms. For these tests, the separation of the analysis windows was done by hand, which introduced the knowledge of the downbeat onsets into the program.

We have measured per window the indigestibility of the set of quantised durations and also the Euclidean distance between the set of quantised durations and the set of performed durations. We have found perfect matches between the score and quantised result in 56 % of the bars from 1955, and in 75 % of the bars from 1981. However, these figures do not imply that the rest of the quantised outcomes is unusable, on the contrary. Because one of our search criteria was the Euclidean distance between the sets of quantised and performed durations, we obtain quantised results that reflect the specific usage of expressive timing by the performer. For example, the common practice of over-dotted note playing when there is only a single dot in the score, is successfully recognised and incorporated in the quantised results. Therefore we have arrived at an additional feature of automated quantisation, which is a usable transcription of expressive timing itself. For example, one can see where Gould has lingered on a particular note value, which typically happens in connection with ornaments. In addition, one can extract information about the rate of ritardando that is performed during the last bar of the *Aria*.

Particularly challenging are those bars that contain long trills combined simultaneously with other ornaments. Earlier tests have shown that a further segmentation into smaller windows can improve the results, i.e., one would find a much closer match with the original score. The same is true for both performances of the last bar 32, which features a long final ritardando.

In chapter 7, we have developed an idea for generating and analysing polyphonic rhythmic sequences from single, composed lines of rhythmic patterns. Edmund Husserl's phenomenology of internal time-consciousness is a seminal work with regard to the philosophy of time. Husserl works with the notion of a continuum of retentions and protentions that originate from the now-point. Because the perceptual present is kept in memory, there are running-off continua of percepts, which start at the now-point. These are called retentions. At any given now-point

the human consciousness samples those retentions, which contain perceptual content from the immediate past. In addition, the consciousness makes projections about when an event is likely to happen next on the basis of the continuous stream of retentions. Processes of learning and entrainment, for example the formation of rhythmic categories and London's MMH, stem from the comparison between past protentions and the present continuum of retentions.

It is very attractive to exploit the above analysis of time consciousness by developing a compositional algorithm for the generation of polyphonic rhythm structures from a monophonic line. We have shown how one can build retentional onset maps, where perceived present note durations are kept in a feedback loop. Such a loop would form a running-off continuum. Depending on the length of the durations involved, there are convergences and divergences of their trajectories down through a number of retentional layers. On each of these layers one would find a different rhythmic pattern, thus we encounter polyphony originating from monophony. An analysis of retentional onset maps leads to the indentification of synchronisation points between onsets on different retentional layers. The synchronisation points are giving information about the nature of the perceived rhythm. Examples include composed accelerando and ritardando effects. Syncopated rhythms are leading to hocket effects, i.e., an interlocking movement of onsets between two or more retentional layers. Onsets on retentional layers accumulate at onset times of metrical pulses. Sometimes such an accumulation can have an amplification effect on a particularly important musical impulse, like we have seen in the Bolero rhythm by Ravel. Rhythmic patterns that are repeated over and over again, an ostinato rhythm, reappear on a lower retentional layer in exactly the same order and with the same durations as the original perceived layer. That means, if metre is regarded as a cyclic pattern of onset points, which feature a high expectancy value for an onset occurring more or less exactly at these points, then the repetitive nature of this cyclic pattern of metre reinforces itself because of the structure of the retentional onset maps. Random rhythms, on the other hand, produce no patterns of synchrony at all between retentional layers, which supports the notion that rhythms with randomly distributed durations are not metrical.

A single line of rhythms generates a polyphony of different voices emerging from a continuum of retentional layers. This leads immediately to numerous possible applications for composition, but also for musicological analysis. The compositional side of these applications is stylistically diverse and can generate a rich material. We also mention the possibility of reversing the generation of retentional layers, such that the perceived layer becomes the last retentional layer in a row. This is a process, which originates in the past of the perceived layer. Finally we mention the possibility of using a non-linearity in the generation of the onset times on retentional layers, which would be very interesting to explore in the future.

Filtered Farey Sequences can model musical rhythms of various cultures and different styles. Furthermore, musical metre, a cyclical rhythmic pattern that is entrained by listeners, can be modelled as well. Rubato and expressive timing introduce continuous change of note durations and for the instantaneous tempo of a piece. Nevertheless it is possible to find the duration classes present in the performance and use a filtered Farey Sequence, combined with indigestibility measurements of the integer ratios involved, in order to arrive at a score transcription in CPN.

The work presented here as a whole and in parts has already proven to be useful in the domains of music performance analysis, musicology, philosophy and composition. We believe that it will continue to evolve and contribute to the knowledge in those fields.

Appendix A

Csound Instrument for Interactive Onset Recording

The instrument listed below triggers a note event using instrument 1300 as the sound generating module. The keyboard's key number pressed together with the timing of that event is written to the file `/tmp/timing.txt`. When the space bar is pressed no sound event is triggered but the event will be recorded in `timing.txt` in order to give the quantisation program the information to start a new analysis window using the next following onset.

```
instr 1

key1 sensekey
if (key1>-1) then
   ktime    times
   printks "%d\\t%f\\n", 0, key1, ktime
   fprintks "/tmp/timing.txt", "%d\\t%f\\n", key1, ktime
   if (key1 != 32) then
      event "i",1300, 0, .2,   .2,  500, 3,  3,  3,  .25, .99,   4,    1, \
      1,   1,  1,  1,  1,  0,  .6,   .05, .1, .4, .2,   .1, .2, .3, .2,  5000, \
      .7,     1,        0
   endif
endif
endin
```

Appendix B

Examples of Quantisation Results Obtained from Commercial Software

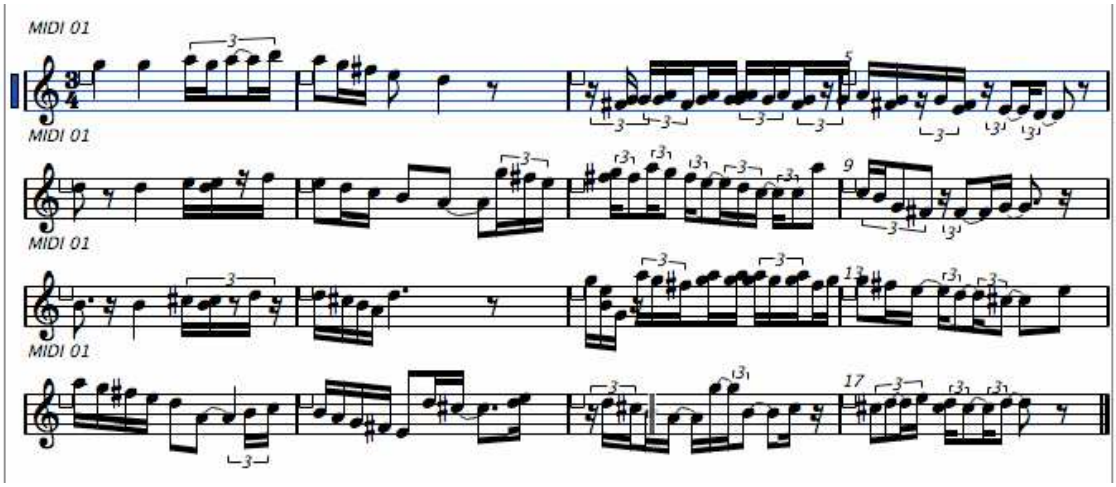


Figure B-1: MIDI performance of the Bach *Aria*, upper voice only, with metronome using Cubase™4.

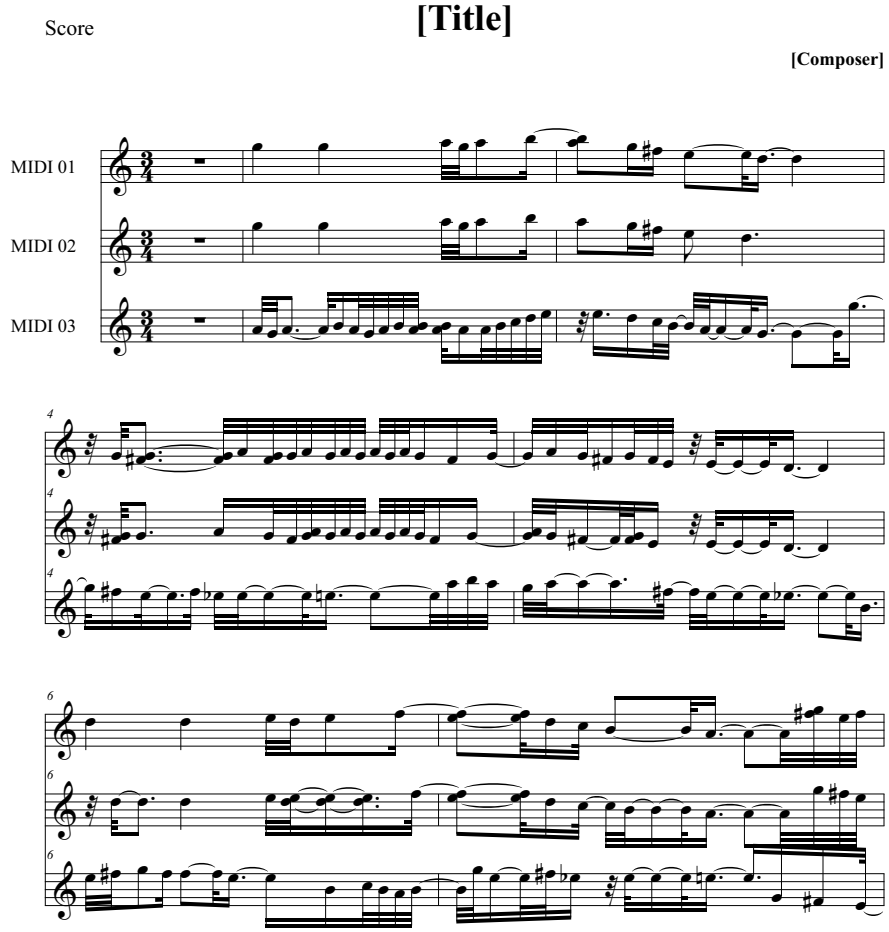


Figure B-2: Rendering of the same MIDI performance as in figure B-1, this time with Finale™.

Bibliography

Agon, C. (1998). *OpenMusic: Un langage de programmation visuelle pour la composition musicale*. PhD thesis, Université Pierre et Marie Curie, Paris 6. Available at `http://recherche.ircam.fr/equipes/repmus/Rapports/CarlosAgon98/index.html` [Accessed August 2006].

Agon, C., Assayag, G., Fineberg, J., and Rueda, C. (1994). Kant: a Critique of Pure Quantification. In *Proceedings of the International Computer Music Conference*, pages 52 – 59, Aarhus, Denmark. ICMA.

al Urmawī, S. A.-D. (1980). *Kitāb al-adwār [Book of Cycles]*. Baghdad. al-Rajab (Ed.).

Arom, S. (1991). *African Polyphony and Polyrhythm*. Cambridge University Press, Cambridge. ISBN 052124160X.

Barlow, C. (1984). *Bus Journey To Parametron (all about* Çoğliuautobüsişletmesi*)*, volume 21-23. Feedback Verlag, Cologne, 2nd edition.

Barlow, C. (1991). Musiquantenlehre. course materials obtained by the author.

Bazzana, K. (1997). *Glenn Gould: The Performer in the Work: A Study in Performance Practice*. Oxford University Press. ISBN 0198166567.

Bencina, R. and Burk, P. (2001). PortAudio - an Open Source Cross Platform Audio API. In *Proceedings of the International Computer Music Conference*, pages 263–266, Havana, Cuba. ICMA.

Berndt, B. C., editor (1994). *Ramanujan's Notebooks*, volume 4. Springer Verlag. page 52-54. ISBN 0-387-94109-6.

Bernstein, L. (1976). *The Unanswered Question*. Harvard University Press, Cambridge. ISBN 0674920015.

Blecksmith, R., McCallum, M., and Selfridge, J. L. (1998). 3-smooth Representations of Integers. *The American Mathematical Monthly*, 105(6):529 – 543. Stable URL: `http://www.jstor.org/stable/2589104` [Accessed April 2010].

Boenn, G. (2005). Development of a Composer's Sketchbook. In *Proceedings of the 3rd International Linux Audio Conference*, pages 73 – 78, Karlsruhe. ZKM. Available at: `http://lac.zkm.de/2005/papers/lac2005_proceedings.pdf`.

Boenn, G. (2007a). Automated Quantisation and Transcription of Ornaments from Audio Recordings. In *Proceedings of the ICMC 2007*, pages 236 – 239, Copenhagen. ICMA, Ann Arbor, Michigan. Jensen K. (Ed.). ISBN: 0-9713192-5-1. Available at: `http://hdl.handle.net/2027/spo.bbp2372.2007.159`.

Boenn, G. (2007b). Composing Rhythms Based Upon Farey Sequences. In *Proceedings of the Digital Music Research Network Conference 2007*, Leeds. Leeds Metropolitan University. 4 pages. Gibson, I., Stansbie, A. and Stavropoulos, N. (Eds.).

Boenn, G. (2008). The Importance of Husserl's Phenomenology of Internal Time-Consciousness for Music Analysis and Composition. In *Proceedings of the ICMC 2008*, Belfast. ICMA, Ann Arbor, Michigan. 4 pages. ISBN 0-9713192-6-X. Available at: `http://hdl.handle.net/2027/spo.bbp2372.2008.157`.

Boenn, G., Brain, M., De Vos, M., and ffitch, J. (2008). Automatic Composition of Melodic and Harmonic Music by Answer Set Programming. In *International Conference on Logic Programming, ICLP08*, volume 4386 of *Lecture Notes in Computer Science*, pages 160–174. Springer Berlin / Heidelberg. ISBN 978-3-540-89981-5.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, 6:145 – 238.

Boulanger, R. (2000). *The Csound Book*. MIT Press, Cambridge. ISBN 0262522616.

Bregman, A. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge. ISBN 0262022974.

Brossier, P. (2006). *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis, Centre for Digital Music Queen Mary, University of London. Available at `http://aubio.org/phd/thesis/brossier06thesis.pdf` [Accessed January 2007].

Brown, J. C. and Smaragdis, P. (2004). Independent component analysis for automatic note extraction from musical trills. *Journal of the Acoustical Society of America*. Available at `http://www.merl.com/reports/docs/TR2004-078.pdf` [Accessed December 2006].

Calkin, N. and Wilf, H. S. (2000). Recounting the Rationals. *The American Mathematical Monthly*, (107):360–363.

Carter, E. (1997). *Collected Essays and Lectures, 1937-1995*. University of Rochester Press. ISBN 1580460259.

Casey, M. and Crawford, T. (2004). Automatic Location and Measurement of Ornaments in Audio Recordings. In *Proceedings of the 5th International Symposium of Music Information Retrieval*, Barcelona. Available at `http://ismir2004.ismir.net/proceedings/p057-page-311-paper252.pdf` [Accessed January 2007].

Celibidache, S. (2008). *Über Musikalische Phänomenologie*. Wissner-Verlag, Augsburg. ISBN 978-3-89639-641-9.

Cemgil, A. T. (2004). *Bayesian Music Transcription*. PhD thesis, Radboud University of Nijmegen. Available at `http://www-sigproc.eng.cam.ac.uk/~atc27/papers/cemgil-thesis.pdf` [Accessed April 2006].

Cemgil, A. T., Desain, P., and Kappen, H. J. (2000). Rhythm Quantization for Transcription. *Computer Music Journal*, 24:2:60–76.

Cemgil, A. T. and Kappen, H. J. (2003). Monte Carlo Methods for Tempo Tracking and Rhythm Quantization. *Journal of Artificial Intelligence Research*, 18:45–81.

Cerrai, P., Pellegrini, C., and Freguglia, P. (2002). *The Application of Mathematics to the Science of Nature: Critical Moments and Aspects*. Springer. ISBN 0306466945.

Chew, G. and Rastall, R. (2001). *The New Grove Dictionary of Music and Musicians*, volume 18, chapter Notation, III, 4-6, pages 140 – 189. Macmillan Publishers Ltd., New York, 2 edition. Sadie S., editor. ISBN 1-56159-239-0.

Clarke, E. F. (1999). *The Psychology of Music*, chapter Rhythm and Timing in Music, pages 473–500. Elsevier, Amsterdam, 2nd edition. Deutsch, D. (Ed.). ISBN 9780122135651.

Collins, N. M. (2006). *Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems*. PhD thesis, University of Cambridge. Available at `http://www.informatics.sussex.ac.uk/users/nc81/research/nickcollinsphd.pdf` [Accessed April 2007].

Crowder, R. G. (1993). *Thinking in Sound: The Cognitive Psychology of Human Audition*, chapter Auditory Memory. Oxford University Press. McAdams, S. and Bigand, E. (Eds.). ISBN 0198522576.

Daniele, J. and Patel, A. D. (2004). The Interplay of Linguistic and Historical Influences on Musical Rhythm in Different Cultures. In *Proceedings of the 8th International Conference on Music Perception and Cognition*, pages 759–762, Adelaide.

Davies, M. (2007). *Towards Automatic Rhythmic Accompaniment*. PhD thesis, Queen Mary University of London, London, UK. Available at `http://www.elec.qmul.ac.uk/digitalmusic/papers/2007/Davies07-phdthesis.pdf` [Accessed May 2008].

de la Motte, D. (1981). *Kontrapunkt*. dtv/Bärenreiter, München. ISBN 342304371.

de Vitry, P. (1964). *Philippe de Vitry: Ars Nova*. Corpus Scriptorum de Musica. Hänssler Verlag, Holzgerlingen.

Desain, P. and Honing, H. (1999). Computational Models of Beat Induction: The Rule-Based Approach. *Journal of New Music Research*, 28(1):29–42.

Desain, P. and Honing, H. (2003). The Formation of Rhythmic Categories and Metric Priming. *Perception*, 32(3):341–365.

Deutsch, D. (1999a). *The Psychology of Music*, chapter Grouping Mechanisms in Music. Elsevier, Amsterdam, 2nd edition. Deutsch, D. (Ed.). ISBN 9780122135651.

Deutsch, D. (1999b). *The Psychology of Music*. Elsevier, Amsterdam, 2nd edition. ISBN 9780122135651.

Didkovsky, N. and Hajdu, G. (2008). MaxScore: Music Notation in Max/MSP. In *Proceedings of ICMC2008*, pages 9–12, Belfast. ICMA.

Dixon, S. (2001). Automatic Extraction of Tempo and Beat from Expressive Performances. *Journal of New Music Research*, 30(1):39–58.

Dupré, M. (1925). *Traité d'Improvisation à l'Orgue*, volume 2 of *Cours Complet d'Improvisation à l'Orgue*. Alphonse Leduc, Paris. Out of print. Available at `http://quod.lib.umich.edu/` [Accessed December 2010].

Efrati, R. R. (1979). *The Interpretation of the Sonatas and Partitas for Solo Violin and the Suites for Solo Cello*. Atlantis, Zürich. 3761105509.

Essl, K. (1996). *Strukturgeneratoren. Algorithmische Komposition in Echzeit*. Beiträge zur Elektronischen Musik. Institut für Elektronische Musik (IEM) an der Universität für Musik und darstellende Kunst in Graz, Graz. Höldrich, R. (Ed.). Available at `http://iem.at/projekte/publications/bem/bem5/` [Accessed December 2011].

Eysenck, M. W. and Keane, M. T. (2005). *Cognitive Psychology*. Psychology Press, 5th edition. ISBN 1841693596.

Fabian, A. (2003). *Bach Performance Practice, 1945-1975*. Aldershot, Hampshire, England. ISBN 0754605493.

Fischer, M., Holland, D., and Rzehulka, B. (1986a). *Gehörgänge. Zur Ästhetik der musikalischen Aufführung und ihrer technischen Reproduktion*, chapter Musik verschwindet. Gespräch der Autoren mit Sergiu Celibidache 1985. Peter Kirchheim. ISBN 3874100162.

Fischer, M., Holland, D., and Rzehulka, B. (1986b). *Gehörgänge. Zur Ästhetik der musikalischen Aufführung und ihrer technischen Reproduktion*, chapter Sergiu Celibidache im Gespräch mit Joachim Matzner. Peter Kirchheim. ISBN 3874100162.

Fraisse, P. (1963). *Psychology of Time*. Harper, New York. ISBN 0837185564.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, 35:1–36.

Fraisse, P. (1999). *The Psychology of Music*, chapter Rhythm and Tempo, pages 149 – 180. Elsevier, Amsterdam, 2nd edition. Deutsch, D. (Ed.). ISBN 9780122135651.

Friberg, A. and Sundberg, J. (1995). Time Discrimination in a Monotonic, Ischronous Sequence. *Journal of the Acoustical Society of America*, 98(5):2524 – 31.

Friberg, A. and Sundström, A. (2002). Swing Ratios and Ensemble Timing in Jazz: Evidence for a Common Rhythmic Pattern. *Music Perception*, 19(3):333–49.

Goto, M. (2001). An Audio-Based Real-Time Beat Tracking System for Music with or without Drum-Sounds. *Journal of New Music Research*, 30(2):159–171. Available at `http://citeseer.ist.psu.edu/goto01audiobased.html` [Accessed April 2006].

Graham, R. L., Knuth, D. E., and Patashnik, O. (1994). *Concrete Mathematics*. Addison-Wesley, Reading, Massachusetts, 2nd edition. ISBN 0201558025.

Gyatso, T. (2005). *The Universe in a Single Atom: The Convergence of Science and Spirituality / H.H. The Dalai Lama XIV*. Morgan Road Books. ISBN 076792066X.

Hainsworth, S. and Macloed, M. (2003). Beat Tracking with Particle Filtering Algorithms. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY. IEEE. Available at: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.2965&rep=rep1&type=pdf` [Accessed January 2007].

Hajdu, G. (1993). Low Energy and Equal Spacing. The Multifactorial Evolution of Tuning Systems. *Interface*, 22:319–333.

Hardy, G. and Wright, E. (1938). *An Introduction to the Theory of Numbers*. Oxford University Press, 4th edition.

Hegel, G. (1975). *Aesthetics : Lectures on Fine Art*. Clarendon Press. trans. Knox, T. M., 2 vols.

Hirsh, I. J. (1959). Auditory Perception of Temporal Order. *Journal of the Acoustical Society of Amercia*, 31(6):759 – 67.

Hirsh, I. J., Monohan, C. B., and al. (1990). Studies in Auditory Timing: 1. Simple Patterns. *Perception and Psychophysics*, 47(3):215–26.

Honing, H. (2001). From Time to Time: The Representation of Timing and Tempo. *Computer Music Journal*, 25(3):50–61.

Honing, H. (2002). Structure and Interpretation of Rhythm and Timing. *Dutch Journal of Music Theory*.

Hoppensteadt, F. C. and Izhikevich, E. M. (1997). *Weakly Connected Neural Networks*, volume 126 of *Applied Mathematical Sciences*. Springer-Verlag, New York. ISBN 0-387-94948-8.

Huron, D. (2006). *Sweet Anticipation. Music and the Psychology of Expectation*. MIT Press. ISBN 0262083450.

Husserl, E. (1966). *The Phenomenology of Internal Time-Consciousness (1883-1917)*. Martinus Nijhoff, The Hague. Boehm, R. (Ed.).

Iverson, P. and Krumhansl, C. L. (1993). Isolating the Dynamic Attributes of Timbre. *Journal of the Acoustical Society of America*, 94(5):2595–2603.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall, Upper Saddle River, NJ. ISBN 013022278X.

James, W. (1890, 1950). *The Principles of Psychology*. Dover, New York. Reprint. ISBN 0486203816.

Kilian, J. (2004). *Inferring Score Level Musical Information From Low-Level Musical Data*. PhD thesis, Technische Universität Darmstadt. Available at `http://www.noteserver.org/kilian/diss-jfk-final.pdf` [Accessed January 2008].

Köhler, W. (1947). *Gestalt Psychology : An Introduction to New Concepts in Modern Psychology*. Mentor, New York, London. ASIN: B0007DEOOS.

Langer, S. (1953). *Feeling and Form : a Theory of Art developed from Philosophy in a New Key*. Routledge and Kegan Paul, London.

Large, E. W. (2001). Periodicity, pattern formation and metric structure. *Journal of New Music Research*.

Large, E. W. (2008). *The Psychology of Time*, chapter Resonating to Musical Rhythm: Theory and Experiment. Emerald, West Yorkshire. Grondin, S. (Ed.). ISBN 9780080469775.

Large, E. W., Almonte, F. V., and Velasco, M. J. (2010). A canonical model for gradient frequency neural networks. *Physica D*, 239(12):905–911. doi:10.1016/j.physd.2009.11.015.

Large, E. W. and Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1):119–159.

Large, E. W. and Kolen, J. F. (1995). *Musical Networks. Parallel Distributed Perception and Performance*, chapter Resonance and the Perception of Musical Meter, pages 65 – 96. MIT Press, Cambridge, Massachusetts. Griffith, Niall and Todd, Peter M. (Eds.). ISBN 0262071819.

Large, E. W. and Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26:1–37.

Lerdahl, F. and Jackendoff, R. (1996). *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Mass., reprint of 1983 edition. ISBN 9780262621076.

London, J. (2004). *Hearing in Time. Psychological Aspects of Musical Meter*. Oxford University Press. ISBN 978-0-19-516081-9.

London, J. (2009). Hearing rhythmic gestures: Moving bodies and embodied minds. World Wide Web electronic publication. Available at `www.people.carleton.edu/~jlondon` [Accessed February 2010.].

Longuet-Higgins, H. C. and Lee, C. S. (1982). The Perception of Musical Rhythms. *Perception*, 11(2):115 128.

Low, E.-L., Grabe, E., and Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language & Speech*, 43:377–401.

MacKay, D. J. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press. ISBN 0521642981.

Markevitch, I. (1983). *Die Sinfonien von Ludwig van Beethoven. Historische, analytische und praktische Studien*. Edition Peters, Leipzig, GDR.

Martelli, A., Ravenscroft, A. M., and Ascher, D., editors (2005). *Python Cookbook*. O'Reilly Media Inc., Sebastopol, California, 2nd edition. ISBN 0596007973.

Martínez, J. M. (2004). Mpeg-7 overview. `http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm`. Acessed 2nd February 2010.

Massaro, D. W. (1970). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology*, 83(1 pt.1):32–39.

Meck, W. H., editor (2003). *Functional and Neural Mechanisms of Interval Timing*. CRC Press, Boca Raton. ISBN 9780203009574.

Messiaen, O. (1995). *Traité de Rhythme, de Couleur et d'Ornithologie*, volume 2. Alphonse Leduc. Publisher Number: AL 28922.

Michon, J. A. (1964). Studies on Subjective Duration. I. Differential Sensitivity in the Perception of Repeated Temporal Intervals. *Acta Psychologica*, 22:441–450.

Michon, J. A. (1978). *Attention and Performance*, volume VII, chapter The Making of the Present: A Tutorial Review. Lawrence Erlbaum, Hillsdale. Raquin, J. (Ed.). Available at: `http://www.jamichon.nl/jam_writings/1976_making_present.pdf` [Accessed September 2010].

Miller, G. H. (1956). The Magical Number Seven, Plus Or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63:81–97.

Moore, F. R. (1990). *Elements of Computer Music*. Prentice Hall. ISBN 0132525526.

Oswald, P. F. (1997). *Glenn Gould. The Ecstasy and Tradegy of Genius*. W. W. Norton & Company. ISBN 0393318478.

Papadelis, G. and Papanikolaou, G. (2004). *The music practitioner : research for the music performer, teacher, and listener*, chapter The Perceptual Space Between and Within Musical Rhythm Categories, pages 117–129. Ashgate, Burlington. Davidson, J.W. (Ed.). ISBN 0754604659.

Parncutt, R. (1994). A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms. *Music Perception*, 11(4):409–464.

Partch, H. (1979). *Genesis of a Music*. Da Capo Press, New York. ISBN 0-306-80106-X.

Patel, A. D. (2008). *Music, Language and the Brain*. Oxford University Press. ISBN 0195123751.

Peeters, G. (2004). A Large Set of Audio Features for Sound Description. Technical report, IRCAM, Paris. Available at: `http://recherche.ircam.fr/equipes/analyse-synthese/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf` [Accessed September 2007].

Pöppel, E. (1972). *The Study of Time*, chapter Oscillations as Possible Basis for Time Perception, pages 219 – 241. Springer-Verlag, Berlin.

Raphael, C. (2001). Automated Rhythm Transcription. In *Proc. Int. Symposium on Music Information Retrieval*, pages 99–107, Bloomington, IN, USA.

Repp, B. H. (1987). The sound of two hands clapping: An exploratory study. *Journal of the Acoustical Society of America*, 81:1100 – 1109.

Repp, B. H. (2002a). Phase correction in sensorimotor synchronization: Nonlinearities in voluntary and involuntary responses to perturbations. *Human Movement Science*, 21(1):1–37.

Repp, B. H. (2002b). Rate limits in sensorimotor synchronization with auditory and visual sequences. In *Meeting on Auditory Perception, Cognition, and Action*, Kansas City, MO.

Repp, B. H. (2003). Phase attraction in sensorimotor synchronization with auditory sequences: Effects of single and periodic distractors on synchronization accuracy. *Journal of Experimental Psychology-Human Perception and Performance*, 29(2):290–309.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin Review*, 12(6):969–992.

Repp, B. H., Windsor, W. L., and al. (2002). Effects of Tempo on the Timing of Simple Musical Rhythms. *Music Perception*, 19(4):565–97.

Right, O. (2001). *The New Grove Dictionary of Music and Musicians*, volume 1, chapter Arab music, I, 1-5, pages 797 – 809. Macmillan Publishers Ltd., New York, 2 edition. Sadie S., editor. ISBN 1-56159-239-0.

Roederer, J. G. (2008). *The Physics and Psychophysics of Music: An Introduction*. Springer Verlag, New York, 4th edition. ISBN 0387094709.

Schaffrath, H. (1995). The Essen Folksong Collection in Kern Format. Computer Database. Menlo Park, CA, Center for Computer Assisted Research in the Humanities. Available at: `http://ota.ahds.ac.uk/headers/1038.xml` [Accessed January 2009].

Schenker, H. (1969). *Five Graphic Music Analyses*. Dover Publications, New York, 2nd edition. ISBN 0486222942.

Schmidt, D. (1999). *Jahrbuch des Staatlichen Instituts fuer Musikforschung Preussischer Kulturbesitz*, chapter Formbildende Tendenzen der Musikalischen Zeit, pages 118–136. Metzler, Stuttgart. ISBN 3476017117.

Schroeder, M. (1991). *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W. H. Freeman and Company. ISBN 0-7167-2136-8.

Sethares, W. A. (2007). *Rhythm and Transforms*. Springer Verlag. ISBN 9781846286391.

Sloane, N. J. A. (2011). The Stern-Brocot Tree. World Wide Web electronic publication. Available at: `http://oeis.org/stern_brocot.html` [Accessed January 2011].

Smalley, D. (1997). Spectromorphology: Explaining Sound-Shapes. *Organised Sound*, 2(2):107–126.

Stockhausen, K. (1988). *Texte zur elektronischen und instrumentalen Musik*. DuMont Buchverlag, Köln.

Stravinsky, I. (1960). *The Rite of Spring*. Columbia Records. Columbia Symphony Orchestra.

Stravinsky, I. (1967). *The Rite of Spring*, volume HPS 638. Boosey & Hawkes, London. Reengraved edition, plate B.&H. 19441. First performance 1913. Revised 1947.

Suchoff, B. (1993). *Béla Bartók Essays*. University of Nebraska Press, Lincoln. ISBN 080326108X.

Takeda, H., Nishimoto, Y., and Sagayama, S. (2003). Automatic transcription of multiphonic MIDI signals. In *Proceedings of the 4th International Conference on Music Infomation Retrieval - ISMIR 2003*, pages 263–264.

Taube, H. (1991). Common Music: A Music Composition Language in Common Lisp and CLOS. *Computer Music Journal*, 15(2):21–32.

Temperley, D. (2007). *Music and Probability*. MIT Press. ISBN 0262201666.

Tenney, J. (1992). *META♮HODOS and META Meta♮Hodos*. Frog Peak Music, 2nd edition. Polansky, L. (Ed.). ISBN 0945996004.

Thakar, M. (1990). *Counterpoint*. Yale University Press. ISBN 0300046383.

Thakar, M. (1999). Tribute to a Teacher. World Wide Web electronic publication. Available at `http://web.archive.org/web/20060129025658/www.celibidache.org/thakar.html` [Accessed February 2010].

The MIDI Manufacturers Association (1996). *The Complete MIDI 1.0 Detailed Specification*. Los Angeles, California. ISBN 0-9728831-0-X.

Toussaint, G. (2004). A Comparison of Rhythmic Similarity Measures. Technical report, School of Computer Science, McGill University. SOCS-TR-2004.6. Available at `http://cgm.cs.mcgill.ca/~godfried/publications/similarity.pdf` [Accessed November 2010].

Verplank, W., Mathews, M., and Shaw, R. (2000). Scanned Synthesis. In *Proceedings of the International Computer Music Conference*, Berlin. ICMA. Available at `http://hdl.handle.net/2027/spo.bbp2372.2000.235` [Accessed December 2010].

Wang, D. (2006). *Computational Auditory Scene Analysis*. Wiley-Interscience, New York. ISBN 0471741091.

Warren, R. M. (1993). *Thinking in Sound*, chapter Perception of Acoustic Sequences: Global Integration versus Temporal Resolution, pages 37 – 68. Oxford University Press. ISBN 0198522576.

Westergaard, P. (1975). *An Introduction to Tonal Theory*. W.W. Norton, New York. ISBN 0393093425.

Wingfield, P. (1999). *Janáček Studies*. Cambridge University Press, Cambridge. ISBN 0521573572.

Woodrow, H. (1932). The effect of rate of sequence upon the accuracy of syncrhonization. *Journal of Experimental Psychology*, 15(4):357 – 79.

Wright, J. and Brandt, E. (2001). System-Level MIDI Performance Testing. In *Proceedings of the International Computer Music Conference*, Havana, Cuba. International Computer Music Association.

Wright, M., Cassidy, R. J., and Zbyszynski, M. F. (2004). Audio and Gesture Latency Measurements on Linux and OSX. In *Proceedings of the International Computer Music Conference*, pages 423–429, Miami. International Computer Music Association. ISBN 0971319227.

Wundt, W. (1911). *Grundzüge der physiologischen Psychologie*. Wilhelm Engelmann, Leipzig.

Xenakis, Y. (1992). *Formalized Music: Thought and Mathematics in Composition*. Pendragon Press, Hillsdale, NY, 2nd edition. ISBN 1576470792.